

Sprechtempo im Sprachvergleich: Eine Untersuchung phonologischer und kultureller Aspekte anhand von Nachrichtensendungen

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
doctor philosophiae
(Dr. phil.)

eingereicht an

der Philosophischen Fakultät II
der Humboldt-Universität zu Berlin

von

Christian Gebhard, M.A.

Präsident der Humboldt-Universität zu Berlin
Prof. Dr. Jan-Hendrik Olbertz

Dekanin der Philosophischen Fakultät II
Prof. Dr. Helga Schwalm

Gutachter: 1. Prof. Dr. Bernd Pompino-Marschall
2. PD Dr. Joachim Christl

Tag der Verteidigung: 18.07.2012

Kurzzusammenfassung

In dieser Arbeit werden verschiedene Sprechtempoparameter in den Sprachen Chinesisch, Deutsch, Englisch, brasilianisches und europäisches Portugiesisch sowie Spanisch zueinander in Bezug gesetzt: Die Maßeinheiten Laute, Silben, Wörter und Morpheme pro Sekunde werden unter Ein- und Ausschluss der Pausenzeit gemessen. Dabei ergeben sich für die germanischen Sprachen niedrige Durchschnittswerte, die romanischen Sprachen liegen im mittleren bis hohen Bereich und das Chinesische weist je nach betrachtetem Parameter mittlere oder hohe Werte auf. Eine Analyse der Pausenstruktur ergibt vergleichbare Pausenanteile, aber einen höheren Anteil interner Pausen in den Sprachen Chinesisch, europäisches Portugiesisch und Spanisch. Weiterhin werden nonverbale Aspekte berücksichtigt und die Größen Gesten, Gesichtsausdrücke und Bewegungen des Kopfes und Oberkörpers pro Minute gemessen. Im brasilianischen Portugiesisch sind Gesten und im Chinesischen Bewegungen des Kopfes und Oberkörpers auffallend häufig. Geschlechtsspezifische Unterschiede sind in allen Bereichen gering und nur im europäischen Portugiesisch statistisch signifikant. Bezüglich des wahrgenommenen Sprechtempos zeigt sich in Experimenten und Umfragen, dass die romanischen Sprachen als schnell bezeichnet werden. Chinesische Hörer schätzen außerdem das Sprechtempo anders ein als deutsche, wobei die Silbenkomplexität für beide Gruppen unterschiedliche Rollen spielt. Daneben bestehen zwischen diesen beiden Gruppen signifikante Unterschiede in der Beurteilung von Sachlichkeit und Vertrauenswürdigkeit in Aufnahmen. Für die untersuchte Sprechsituation wird davon ausgegangen, dass in verschiedenen Sprachgemeinschaften die Verwendung unterschiedlicher sprachlicher (verbaler wie nonverbaler) Mittel bevorzugt wird. Das Sprechtempo wird damit als eine von phonetisch-phonologischen Faktoren, aber auch von in gesellschaftlichen Gruppen vorhandenen Tendenzen im Sprachgebrauch beeinflusste Größe verstanden.

Schlagworte: Sprechtempo, kontrastive Linguistik, Kulturvergleich, nonverbale Kommunikation

English abstract

This doctoral thesis is a cross-linguistic analysis of speech tempo, comparing Chinese, English, German, Brazilian and European Portuguese, and Spanish. There are several parameters of speech tempo under analysis: sounds, syllables, words, and morphemes per second. The Germanic languages show lower values, whereas Romance languages tend to have medium and high values, and Chinese appears to have different, both medium and high values depending on what element is counted. An analysis of pause structure renders comparable percentages of time spent on silence, yet the proportion of internal pauses is considerably higher in Chinese, European Portuguese and Spanish. Furthermore, nonverbal aspects of speech are considered and data are collected for the measures gestures, facial expressions and movements of the head and torso per minute. Gestures are remarkably more frequent in Brazilian Portuguese, and movements of the head/torso are highly frequent in Chinese. Differences between genders are generally small, only within European Portuguese some statistically significant differences can be observed. Surveys on stereotypes and perception experiments show that the Romance languages are generally perceived as spoken more rapidly. The listeners' native language appears to be a factor of perceived speech tempo: Chinese listeners' responses are significantly different from German listeners' responses. Syllable complexity plays different roles for these two groups of listeners. They also show significant differences when judging objectivity and trustworthiness in recordings. Considering the particular communication situation analyzed here, it is pointed out that in different speech communities, different verbal and nonverbal behavior is preferred in certain contexts. Speech tempo is influenced by phonetic and phonological factors, as well as tendencies towards a certain communicative style that can be observed within a given society.

Keywords: Speech tempo, contrastive linguistics, cross-culture comparison, nonverbal behavior

INHALT

KURZZUSAMMENFASSUNG	II
ENGLISH ABSTRACT	III
ABBILDUNGSVERZEICHNIS	VIII
TABELLENVERZEICHNIS	X

ERSTER TEIL

1 EINLEITUNG	1
2 GRUNDLEGENDE BEGRIFFE UND ÜBERLEGUNGEN	7
2.1 Definitionen	7
2.1.1 Sprechtempo	7
2.1.1.1 Sprech- und Artikulationsgeschwindigkeit	8
2.1.1.2 Zählbare sprachliche Elemente	9
2.1.2 Zeit	15
2.1.3 Definition von „schnell“ und „langsam“	17
2.2 Sprachvergleich: schnelle vs. langsame Sprachen?	20
2.3 Faktoren des Sprechtempos	23
2.3.1 Faktoren, die das lokale Sprechtempo beeinflussen können	24
2.3.2 Faktoren, die das globale Sprechtempo beeinflussen können	30
2.4 Pausen	39
2.4.1 Arten von Pausen	39
2.4.2 Zur Wahrnehmung von Pausen	43
2.4.3 Pausen im Sprachvergleich	44
2.4.4 Zusammenhänge von Sprechgeschwindigkeit, Artikulationsgeschwindigkeit und Pausen	47
2.5 Zusammenfassung	48
3 PHONETISCH-PHONOLOGISCHE DETERMINANTEN DES SPRECHTEMPOS	50
3.1 Einleitung	50
3.2 Bisherige Forschungsergebnisse sprachvergleichender Untersuchungen zum Sprechtempo	51
3.3 Koartikulation im Sprachvergleich	54
3.4 Sprechtempo und Rhythmustypologien	57
3.4.1 Definition von Rhythmus und weiteren relevanten Begriffen	57
3.4.1.1 Rhythmus	57
3.4.1.2 Rhythmus und Sprache	60
3.4.2 Rhythmus im phonetischen Sprachsignal: Rhythmusmesswerte	61

3.4.3	<i>Entrainment</i>	68
3.4.4	Rhythmus und Tempo	70
3.5	Zusammenfassung	76
4	EMPIRISCHE STUDIE ZUM SPRECHTEMPO	77
4.1	Einleitung	77
4.2	Phonetisch-phonologische Vorstellung der Einzelsprachen: Brasilianisches Portugiesisch, Chinesisch, Deutsch, Englisch, Europäisches Portugiesisch und Spanisch	78
4.3	Material	97
4.4	Methodik	99
4.4.1	Aufnahme und Transkription	99
4.4.2	Definitionen der ausgezählten sprachlichen Elemente	100
4.4.2.1	Definition von Lauten als zählbare sprachliche Elemente	100
4.4.2.2	Definition von Silben als zählbare sprachliche Elemente	102
4.4.2.3	Definition von Morphemen als zählbare sprachliche Elemente	104
4.4.2.4	Definition von Wörtern als zählbare sprachliche Elemente	104
4.4.3	Transkriptionsbeispiel	105
4.4.4	Pausenanalyse	105
4.5	Ergebnisse	107
4.5.1	Brasilianisch	109
4.5.2	Chinesisch	110
4.5.3	Deutsch	111
4.5.4	Englisch	113
4.5.5	Portugiesisch	113
4.5.6	Spanisch	116
4.5.7	Vergleich	118
4.5.7.1	Tempo	118
4.5.7.2	Rhythmus	125
4.5.7.3	Informationsdichte und Form der Kodierung	126
4.5.7.4	Pausenstruktur	128
4.5.7.5	Chinesisch und Deutsch: Ein lokaler Vergleich	137
4.6	Zusammenfassung	138
5	DIE WAHRNEHMUNG DES SPRECHTEMPOS	141
5.1	Einleitung	141
5.2	Definition	141
5.3	Muttersprache vs. Fremdsprache	142
5.4	Faktoren, die das wahrgenommene Sprechtempo beeinflussen	146

VI

5.4.1	Grundfrequenz	146
5.4.2	Pausen	147
5.4.3	Artikulatorische Genauigkeit	148
5.4.4	Silbenstrukturen	149
5.4.5	Erwartungshaltung (Stereotype)	149
5.4.6	Weitere Faktoren	150
5.5	Empirische Untersuchung zum wahrgenommenen Sprechtempo	151
5.5.1	Vorläufige Untersuchung	151
5.5.2	Wahrnehmungsexperiment zur Bedeutung der Silbenstruktur	153
5.5.2.1	Methode	153
5.5.2.2	Teilnehmer	153
5.5.2.3	Stimuli	154
5.5.2.4	Ergebnisse	155
5.6	Zusammenfassung	157

ZWEITER TEIL

6	KULTURELLE ASPEKTE DES SPRECHTEMPOS	159
6.1	Einleitung	159
6.2	Definition des Kulturbegriffs	159
6.3	Bisherige Untersuchungen	161
6.4	Vergleich der Ergebnisse aus Kapitel 4 mit kulturellen Dimensionen	165
6.5	Umfrage zu kulturellen Stereotypen	170
6.6	Wahrnehmungsexperiment zur Bewertung von Sachlichkeit und Vertrauenswürdigkeit	172
6.6.1	Methode	173
6.6.2	Teilnehmer	173
6.6.3	Stimuli	173
6.6.4	Ergebnisse und Diskussion	174
6.7	Zusammenfassung	176
7	NONVERBALE KOMMUNIKATION IM SPRACHVERGLEICH	178
7.1	Einleitung	178
7.2	Definition und Bedeutung nonverbaler Kommunikation	178
7.2.1	Definition	178
7.2.2	Bedeutung nonverbaler Kommunikation	179
7.2.2.1	Anteil nonverbaler Kommunikation an sprachlicher Kommunikation	179
7.2.2.2	Zusammenhang von Sprache und Gestik / Kinesik	180
7.3	Teilbereiche und Klassifizierungen nonverbaler Kommunikation	182

VII

7.3.1	Teilbereiche nonverbaler Kommunikation	182
7.3.2	Klassifizierungen von Gesten	185
7.4	Kulturelle Unterschiede in nonverbaler Kommunikation	187
7.4.1	Gesichtsausdrücke: Universalien und kulturelle Normierungen	187
7.4.2	Gestik: Qualitative Unterschiede	189
7.4.2.1	Embleme	189
7.4.2.1.1	Inhalt, der mittels verschiedener Zeichen ausgedrückt wird	190
7.4.2.1.2	Zeichen, die unterschiedliche Bedeutungen ausdrücken	191
7.4.2.1.3	Zeichen, die keine Entsprechungen finden	191
7.4.2.2	Gestenraum	192
7.4.3	Gestik und Mimik: Quantitative Unterschiede und Erklärungsansätze	192
7.5	Zusammenfassung	198
8	EMPIRISCHE UNTERSUCHUNG ZUR NONVERBALEN KOMMUNIKATION	199
8.1	Einleitung	199
8.2	Methode	200
8.3	Material	201
8.4	Ergebnisse	204
8.4.1	Quantitative Untersuchung	204
8.4.1.1	Faktor Sprache	204
8.4.1.2	Faktor Geschlecht	206
8.4.1.3	Faktor Register	207
8.4.2	Qualitative Untersuchung	209
8.4.2.1	Gestik	209
8.4.2.1.1	Dauer der Gesten	209
8.4.2.1.2	Gestenraum	210
8.4.2.1.3	Motivierte Gesten	211
8.4.2.2	Persönlicher Bezug	214
8.4.2.3	Mimik	215
8.4.2.4	Kinesik	217
8.4.2.5	Weitere Beobachtungen	220
8.4.3	Diskussion	222
8.5	Zusammenfassung	225
9	ZUSAMMENFASSUNG DER ERGEBNISSE	227
	LITERATURVERZEICHNIS	235
	ANHANG	264

Abbildungsverzeichnis

Abbildung 2.1: SZ inkl. markiertem Abschnitt der Stille	8
Abbildung 4.1: BP „pneus“ mit Vokalepenthese [pi'neʊs]	80
Abbildung 4.2: Vortonige Nullrealisierung von /e/ in EP "se festejou"	92
Abbildung 4.3: Kurze stille Pause	106
Abbildung 4.4: Phase des totalen Verschlusses	106
Abbildung 4.5: Anteil der SZ der untersuchten Sprachen am Gesamtkorpus	108
Abbildung 4.6: Komplexe Konsonantenhäufungen ohne Elision im Deutschen	112
Abbildung 4.7: Geschlechtsspezifische Unterschiede im EP (AG in Lauten/s)	115
Abbildung 4.8: Geschlechtsspezifische Unterschiede im EP (AG in Silben/s)	115
Abbildung 4.9: Geschlechtsspezifische Unterschiede im EP (Pausenanteil)	116
Abbildung 4.10: Negative VOT nach Pause in es. „retoma“ [nre'toma]	117
Abbildung 4.11: Stark reduzierte Aussprache von es. /'xweyos/	118
Abbildung 4.12: Lautgeschwindigkeiten (AG) der untersuchten Sprachen	119
Abbildung 4.13: Silbengeschwindigkeiten (AG) der untersuchten Sprachen	120
Abbildung 4.14: Wortgeschwindigkeiten der untersuchten Sprachen	121
Abbildung 4.15: SGs der untersuchten Sprachen im Vergleich	122
Abbildung 4.16: Morphemgeschwindigkeiten der untersuchten Sprachen	127
Abbildung 4.17: Zusammenhang zwischen Lauten/Silbe und Silben/Wort	127
Abbildung 4.18: Pausenanteil in den untersuchten Sprachen (in %)	129
Abbildung 4.19: Anteil interner Pausen an allen Pausen (in %)	130
Abbildung 4.20: Durchschnittliche Pausendauer (in s)	131
Abbildung 4.21: Durchschnittliche Pausenhäufigkeit (Abstand zwischen Pausen in s)	131
Abbildung 4.22: f0-Deklination in syntaktisch unvollständigem Satz bei spanischem Nachrichtensprecher	134
Abbildung 5.1: Laut- und Silbengeschwindigkeiten der Stimuli des Wahrnehmungsexperiments	154
Abbildung 8.1: Häufigkeit der Gesten/min in den untersuchten Sprachen	204
Abbildung 8.2: Häufigkeit der Gesichtsausdrücke/min in den untersuchten Sprachen	205
Abbildung 8.3: Häufigkeit der Bewegungen des Oberkörpers und/oder Kopfes/min in den untersuchten Sprachen	205
Abbildung 8.4: Verwendung von Gestik in den unterschiedlichen Registerebenen	208
Abbildung 8.5: Gestensequenz eines Nachrichtensprechers (BP)	210
Abbildung 8.6: Beanspruchter Raum in Ruheposition spanischer Nachrichtensprecherin	211
Abbildung 8.7: Beanspruchter Raum in Ruheposition deutscher Nachrichtensprecherin	211
Abbildung 8.8: Räumlich-zeitliche motivierte Geste von brasilianischem Nachrichtensprecher	212
Abbildung 8.9: Von brasilianischem Nachrichtensprecher verwendetes Emblem	212
Abbildung 8.10: Brasilianischer Nachrichtensprecher stellt den Begriff gestisch dar	212
Abbildung 8.11: Deutsche Sprecherin untermalt den Ausdruck „verschmolzen“	212
Abbildung 8.12: Chinesischer Sprecher strukturiert Aufzählung	212

Abbildung 8.13: Brasilianische Nachrichtensprecherin betont Zahl _____	213
Abbildung 8.14: Verabschiedung einer brasilianischen Nachrichtensprecherin _____	215
Abbildung 8.15: Verabschiedung einer deutschen Nachrichtensprecherin _____	215
Abbildung 8.16: Direkter Bezug eines brasilianischen Nachrichtensprechers zum Zuschauer__	215
Abbildung 8.17: Direkter Bezug einer deutschen Nachrichtensprecherin zum Zuschauer _____	215
Abbildung 8.18: Kamerafahrt in spanischer Nachrichtensendung_____	218
Abbildung 8.19: Kamerafahrt in portugiesischer Nachrichtensendung _____	219
Abbildung 8.20: Proxemisches Verhalten bei Sprecherwechsel in brasilianischer Nachrichtensendung _____	222
Abbildung 10.1: Grafische Oberfläche des Wahrnehmungsexperiments (5.5) mit chinesischen Teilnehmern _____	272

Tabellenverzeichnis

Tabelle 3.1: Übersicht über RMs	63
Tabelle 4.1: Konsonantensystem des BP	78
Tabelle 4.2: Vokalsystem des BP	78
Tabelle 4.3: Häufigkeit der Konsonanten des BP im gesprochenen Text	79
Tabelle 4.4: Silbenstrukturen im BP	80
Tabelle 4.5: Konsonantensystem des Chinesischen	81
Tabelle 4.6: Vokalsystem des Chinesischen	82
Tabelle 4.7: Häufigkeit chinesischer Konsonanten im gesprochenen Text	82
Tabelle 4.8: Silbentypen des Chinesischen	84
Tabelle 4.9: Konsonantensystem des Deutschen	86
Tabelle 4.10: Vokalsystem des Deutschen	86
Tabelle 4.11: Häufigkeit deutscher Konsonanten im gesprochenen Text	87
Tabelle 4.12: Silbentypen im Deutschen	88
Tabelle 4.13: Konsonantensystem des Englischen	89
Tabelle 4.14: Vokalsystem des Englischen	89
Tabelle 4.15: Häufigkeit der englischen Konsonanten im gesprochenen Text	89
Tabelle 4.16: Silbentypen im Englischen	90
Tabelle 4.17: Konsonantensystem des EP	91
Tabelle 4.18: Vokalsystem des EP	91
Tabelle 4.19: Häufigkeit der Konsonanten des EP im gesprochenen Text	92
Tabelle 4.20: Silbentypen im EP	93
Tabelle 4.21: Konsonantensystem des Spanischen	94
Tabelle 4.22: Vokalsystem des Spanischen	95
Tabelle 4.23: Häufigkeit spanischer Konsonanten im gesprochenen Text	95
Tabelle 4.24: Silbentypen im Spanischen	96
Tabelle 4.25: Gerundete SZs (in s) der aufgenommenen Sprecher	108
Tabelle 4.26: Sprechtempomesswerte - Ergebnisse für BP	110
Tabelle 4.27: Sprechtempomesswerte - Ergebnisse für Chinesisch	111
Tabelle 4.28: Sprechtempomesswerte - Ergebnisse für Deutsch	112
Tabelle 4.29: Sprechtempomesswerte - Ergebnisse für Englisch	113
Tabelle 4.30: Sprechtempomesswerte - Ergebnisse für EP	114
Tabelle 4.31: Sprechtempomesswerte - Ergebnisse für Spanisch	117
Tabelle 4.32: Korrelationen zwischen Sprechtempomesswerten und segmentaler Beschaffenheit des Sprachmaterials (Sprachvergleich)	124
Tabelle 4.33: Komplexität verschiedener sprachlicher Elemente der untersuchten Sprachen	128
Tabelle 4.34: Pausenanteil, Anteil interner Pausen, Pausendauer und Dauer artikulatorischer Phrasen der einzelnen Sprachen	132
Tabelle 4.35: Korrelationen zwischen SGs und Pausen	133

Tabelle 4.36: Mittelwerte für die Konsonanten(-häufungen) [ts ^h], [t ^h] und [s] im Chinesischen und Deutschen	137
Tabelle 5.1: Beurteilung der SG der Sprachen BP, DE, EN und ES von zehn Muttersprachlern des Chinesischen	152
Tabelle 5.2: Korrelationsmatrix des Wahrnehmungsexperiments: gemittelte Werte auf Wahrnehmungsskala und phonetische Beschaffenheit der Stimuli	155
Tabelle 6.1: Korrelationen zwischen Sprechtempo und kulturellen Dimensionen	167
Tabelle 6.2: Ergebnisse der Befragungen zur typischen SG anderer Nationalitäten	172
Tabelle 6.3: Beschaffenheit der Aufnahmen im Versuch zur wahrgenommenen Sachlichkeit und Vertrauenswürdigkeit	174
Tabelle 6.4: Signifikante Unterschiede in der Beurteilung von Sachlichkeit bzw. Vertrauenswürdigkeit	175
Tabelle 6.5: Zusammenhänge zwischen gemittelter wahrgenommener Sachlichkeit und Vertrauenswürdigkeit mit Sprechtempomesswerten, f0-Wechseln und Pausenhäufigkeit	176
Tabelle 8.1: Auf die Verwendung nonverbaler Kommunikationsmittel hin untersuchte Nachrichtensendungen	203
Tabelle 8.2: Mittelwerte und Standardabweichungen für produzierte Gesten, mimische Ausdrücke und Veränderungen der Position des Kopfes und/oder Oberkörpers/min	206
Tabelle 8.3: Unterschiede zwischen den Geschlechtern: produzierte Gesten, mimische Ausdrücke und Veränderungen der Position des Kopfes und/oder Oberkörpers/min	207
Tabelle 8.4: Korrelationen von Gestik, Mimik und Bewegungen mit Registerebene	208
Tabelle 8.5: Zusammenhang zwischen Sprechtempo und Häufigkeit nonverbaler Kommunikationsmittel	225
Tabelle 10.1: Rohdaten der Sprechtempomesswerte: Mittelwerte pro Sprecher	265
Tabelle 10.2: Einfaktorielle Varianzanalyse: F-Werte (Variablen: Tempomesswerte, Faktor: Sprache)	268
Tabelle 10.3: Einfaktorielle Varianzanalyse: Post-Hoc-Vergleiche nach Scheffé (Variablen: Tempomesswerte, Faktor: Sprache)	268
Tabelle 10.4: Bewertung des Sprechtempos der Stimuli durch deutsche und chinesische Muttersprachler	272
Tabelle 10.5: Lineares Regressionsmodell zur Erklärung der gemittelten wSG deutscher Muttersprachler	275
Tabelle 10.6: Lineares Regressionmodell zur Erklärung der gemittelten wSG chinesischer Muttersprachler	275
Tabelle 10.7: Rohdaten der Häufigkeit der Verwendung nonverbaler Kommunikationsmittel	276
Tabelle 10.8: Einfaktorielle Varianzanalyse: F-Werte (Variablen: Häufigkeiten der Verwendung nonverbaler Kommunikationsmittel, Faktor: Sprache)	277
Tabelle 10.9: Einfaktorielle Varianzanalyse: Post-Hoc-Vergleiche nach Scheffé (Variablen: Häufigkeiten der Verwendung nonverbaler Kommunikationsmittel, Faktor: Sprache)	278

1 Einleitung

Die Begriffe Tempo, Geschwindigkeit und Schnelligkeit bezeichnen ganz allgemein betrachtet, wie schnell ein Vorgang abläuft und sollen vorerst synonym gebraucht werden. Darunter kann verstanden werden, wie viel in einem Zeitraum geschieht, welcher Weg in einem Zeitraum zurückgelegt wird, oder wie häufig sich ein Ereignis wiederholt. Für sprachliches Verhalten können unterschiedliche Arten von Schnelligkeit untersucht werden, z.B. in welchem Zeitraum ein bestimmter Weg von den Artikulationsorganen zurückgelegt wird, also mit welchem Tempo (ohne Berücksichtigung der Richtung des sich bewegenden Körpers) artikuliert wird. Etwas vereinfacht kann, ohne Berücksichtigung des Weges, die Anzahl sprachlicher Elemente pro Zeiteinheit gezählt werden, wobei hier zunächst diskutiert werden muss, welche sprachlichen Elemente zu zählen sind. Diese können artikulatorisch, aber auch inhaltlich definiert werden. Neben diesen lautsprachlichen Einheiten kommen auch nonverbale kommunikative Zeichen in Frage. So kann untersucht werden, wie viele gestische Zeichen in einem Zeitraum produziert werden. Da es sich dabei, wenn lautsprachliche Kommunikation untersucht wird, um sprachbegleitende Zeichen handelt, die prinzipiell fakultativen Charakter aufweisen, ist eher von der Häufigkeit als der Geschwindigkeit zu sprechen. In der Alltagssprache werden Sprecher als schnell oder langsam bezeichnet, aber für Urteile hierüber können meist keine objektiv gemessenen Werte als Basis genannt werden.

Nach einem weiteren Verständnis von Sprechtempo kann untersucht werden, wie viel Inhalt notwendig ist, um eine bestimmte Nachricht mitzuteilen. Diese Auffassung kann als „Kommunikationsgeschwindigkeit“ bezeichnet werden und wird umgangssprachlich oft damit umschrieben, „wie schnell jemand auf den Punkt kommt“. Die genannten Auffassungen von dem, was unter dem Begriff „Sprechtempo“ zusammengefasst werden soll, werden in dieser Arbeit zu unterschiedlichen Teilen berücksichtigt.

Nicht nur Individuen, sondern auch Gruppen von Sprechern werden als schnell oder langsam bezeichnet. Von Fremdsprachenlernern werden häufig Urteile bezüglich der Schnelligkeit der Zielsprache gefällt. Viele sind sich hierbei einer möglichen Überschätzung bewusst und vergleichen die Schnelligkeit in der Fremdsprache mit der eigenen Sprechgewohnheit, wobei auch dann bestimmte Sprachen als schneller, andere als

langsamer empfunden werden. *Klingen* einige Sprachen z.B. auf Grund lautlicher oder rhythmischer Merkmale schneller als andere?

Untersuchungen liefern widersprüchliche Antworten auf die Frage, ob Sprachen tatsächlich unterschiedlich schnell *gesprochen* werden. Die Beantwortung dieser Frage wird durch eine teilweise zu einseitige Definition von Schnelligkeit in Bezug auf Sprache erschwert: In vielen Studien werden Laute, Silben oder Wörter pro Zeiteinheit gezählt. Anhand dieser Daten werden (oft nur zwei) Sprachen miteinander verglichen, in denen die genannten sprachlichen Elemente sehr unterschiedliche Längen aufweisen können. Auch wenn das Sprechtempo im Schatten der Rhythmusklassenforschung betrachtet wird, ergeben sich kaum neue Erkenntnisse über die zeitliche Organisation der einzelsprachlichen Produktion. Daher wird in der vorliegenden Arbeit ein Sprachvergleich in einer normierten Sprechsituation unter Berücksichtigung verschiedener Parameter des Sprechtempos angestellt. Dabei soll nicht nur die Frage beantwortet werden, ob unterschiedliche Tempi in Abhängigkeit des Faktors Sprache festgestellt werden können, sondern inwiefern einerseits phonetisch-phonologische Gegebenheiten, andererseits kulturelle Unterschiede bestimmte Sprechgeschwindigkeiten auch innerhalb einer Sprache erklären können.

Anhand des Vergleichs mehrerer Sprachen kann festgestellt werden, ob z.B. die Silbenkomplexität, die auch als Unterscheidungsmerkmal zwischen phonologischen Sprachtypen dient, oder die Häufigkeit bestimmter intrinsisch längerer/kürzerer Laute im phonologischen Lautinventar einer Sprache das Sprechtempo beeinflussen. Vorstellbar ist, dass aufgrund geringerer artikulatorischer Flexibilität im Vergleich mit Vokalen Konsonanten ab einer bestimmten Häufigkeit eine das Sprechtempo reduzierende Wirkung haben.

Weiterhin wird eine kulturelle Perspektive eingenommen und gefragt, ob gesellschaftlich festgelegte Präferenzen für eine bestimmte Sprechgeschwindigkeit plausibel sind und welche Urteile mit verschiedenen Tempi einhergehen. Ergebnisse der interkulturellen Psychologie zeigen, dass kulturelle Normen einen Einfluss nicht nur auf die Persönlichkeitswahrnehmung, sondern auch auf die verbale und nonverbale Kommunikation haben. Es ist durchaus vorstellbar, dass in manchen Sprachgemeinschaften langsames, in anderen hohes Sprechtempo positiv bewertet wird.

Eng mit dem Sprechtempo verbunden ist die Art, wie Pausen bei der Sprachproduktion eingesetzt werden. Eine qualitative und quantitative Untersuchung von Pausen ist notwendig, um festzustellen, ob im Sprachvergleich Unterschiede auffallen, die nicht auf

die linguistische Beschaffenheit des Sprachmaterials zurückzuführen sind und somit von außersprachlichen, in Sprachgemeinschaften festgelegten Normen abhängig sind.

Als Untersuchungsgegenstand dienen Nachrichtensendungen in Form von Webcasts aus Brasilien, China, Deutschland, Großbritannien, Portugal und Spanien. Die untersuchten Sprachen wurden so gewählt, dass germanische mit romanischen Sprachen verglichen werden können. Dieser Vergleich ist deshalb von Interesse, weil in der Literatur Hinweise gegeben wurden, dass romanische Sprachen im Vergleich mit germanischen schneller gesprochen werden, allerdings ohne überzeugende Beweise, die auf empirischen Untersuchungen gründen. Der Vergleich des brasilianischen mit dem europäischen Portugiesisch lässt außerdem einen Aufschluss darüber zu, ob sich sehr eng verwandte Sprachen auch in Bezug auf das Sprechtempo ähnlich sind. Darüber hinaus wird zu diesem Zweck eine außereuropäische Sprache zum Vergleich herangezogen. Zu den Bezeichnungen der Einzelsprachen sei hier angemerkt, dass brasilianisches Portugiesisch als Sprache behandelt und nicht mit der europäischen Varietät des Portugiesischen als ein sprachliches System bezeichnet wird. Diese Anschauung fundiert in der Berücksichtigung phonologischer Unterschiede zwischen den beiden Sprachen (die auch als Varietäten bezeichnet werden), die hier als Faktor des Sprechtempos untersucht werden. Weiterhin wird auch von kulturellen Unterschieden ausgegangen, die trotz Bestrebungen zur Vereinheitlichung des Schriftbildes zusammen mit der geographischen Distanz zwischen den beiden Ländern nicht außer Acht gelassen werden dürfen. Im Fall des Chinesischen wird hier nicht zwischen verschiedenen Dialekten und (gegenseitig nicht verständlichen) Sprachen unterschieden und ausschließlich die als „Mandarin“ oder „Hochchinesisch“ (普通话 pǔtōnghuà, de. etwa „Allgemeinsprache“) bezeichnete Varietät untersucht. Hierbei wird einheitlich die Bezeichnung „Chinesisch“ verwendet. Im Folgenden wird daher von den Sprachen brasilianisches Portugiesisch (BP), Chinesisch (CN), Deutsch (DE), Englisch (EN), europäisches Portugiesisch (EP) und Spanisch (ES) die Rede sein. Diese Sprachen werden auf unterschiedliche Eigenschaften hin analysiert und die Ergebnisse in verschiedenen Kapiteln im Anschluss an theoretische Überlegungen und entsprechende Zusammenfassungen der bisherigen relevanten Forschungsliteratur präsentiert und diskutiert. Es wird wie folgt vorgegangen:

Das zweite Kapitel beginnt mit der Definition grundlegender Begriffe und Konzepte. Darin wird außerdem die Wahl der in der sprachvergleichenden Betrachtung des Sprechtempos zu zählenden sprachlichen Elemente diskutiert. Von dieser Wahl hängt das Ergebnis empirischer Untersuchungen der Schnelligkeit bei der Produktion gesprochener Sprache ab, wie anhand von Beispielen verdeutlicht wird. Es beinhaltet weiterhin in einem

Literaturüberblick eine ergänzende Übersicht zu den bisher untersuchten Faktoren des Sprechtempos. Im dritten Kapitel wird auf phonetisch-phonologische Aspekte des Sprechtempos und relevante Forschungsergebnisse eingegangen, die u.a. die Themen Koartikulation und Sprachrhythmus und deren Bezug zum Tempo berühren. Im vierten Kapitel werden dann die Ergebnisse der empirischen Studie bezüglich der verschiedenen Aspekte des Sprechtempos dargestellt. Dabei wird neben den erhobenen Werten für die Sprechgeschwindigkeit auf die Pausenstruktur und sprachrhythmische Unterschiede eingegangen. Die Wahrnehmung des Sprechtempos ist Gegenstand des fünften Kapitels, in dem bisher untersuchte Faktoren zusammengefasst werden und anschließend eine empirische Untersuchung zum Einfluss der Muttersprache auf das wahrgenommene Tempo präsentiert wird.

Im zweiten Teil der Arbeit werden kulturelle Aspekte untersucht. Im in den soziokulturell orientierten Teil der Arbeit einleitenden sechsten Kapitel, in dem zunächst grundlegende Begriffe wie z.B. „Kultur“ definiert und weitere, nicht phonetisch orientierte Arbeiten zum Sprechtempo vorgestellt werden, werden auch die Ergebnisse der phonetisch-phonologisch orientierten Studie mit Ergebnissen von Untersuchungen verglichen, in denen Gesellschaften anhand sogenannter kultureller Dimensionen voneinander unterschieden werden. Darüber hinaus werden eine eigene Umfrage zu kulturellen Stereotypen und ein Wahrnehmungsexperiment, das die Einschätzung von Sachlichkeit und Vertrauenswürdigkeit in den untersuchten Nachrichtensendungen zum Gegenstand hat, vorgestellt. Anschließend werden nonverbale Komponenten der Sprache untersucht. Das siebte Kapitel beschäftigt sich mit nonverbaler Kommunikation zunächst im Allgemeinen, dann im Sprachvergleich. Die Ergebnisse der Auswertung der Nachrichtensendungen mit Bezug auf die Verwendung dieser Kommunikationsmittel werden anschließend im achten Kapitel vorgestellt. Welche Bedeutung phonetisch-phonologische und kulturelle Gegebenheiten für das Sprechtempo haben, wird in den zusammenfassenden Bemerkungen des neunten Kapitels dargestellt.

Zur Formulierung soll angemerkt werden, dass der flüssigeren Lesbarkeit halber grammatische Maskulina als Personenbezeichnungen verwendet werden, die sich aber ausdrücklich auf Personen weiblichen und männlichen Geschlechts beziehen.

Für Sprachbezeichnungen werden im durchlaufenden Text die im ISO Standard 639-1 festgelegten Abkürzungen verwendet: cn. (chinesisch), de. (deutsch), en. (englisch), fr. (französisch), ja. (japanisch), pt. (portugiesisch, wobei, wenn nötig, zwischen brasilianischem Portugiesisch, bp., und europäischem Portugiesisch, ep., unterschieden

wird), es. (spanisch) und tr. (türkisch). Beispiele aus dem Chinesischen werden in vereinfachten Schriftzeichen mit Pinyin, dem in der Volksrepublik China üblichen Transskriptionssystem mit lateinischen Buchstaben, und deutscher Übersetzung angegeben.

Erster Teil

2 Grundlegende Begriffe und Überlegungen

Dieses Kapitel beginnt mit der Definition von „Sprechtempo“ und einigen relevanten Konzepten wie dem Zeitbegriff sowie der Vorstellung dessen, was im Kontext gesprochener Sprache als „schnell“ oder „langsam“ bezeichnet werden kann. Weiterhin werden in einem Literaturüberblick mögliche Einflüsse auf das Sprechtempo beschrieben und Erkenntnisse der Pausenforschung vorgestellt. Dabei wird auf einige Untersuchungen eingegangen, ohne dass dabei Anspruch auf Vollständigkeit erhoben wird, da es sich um einleitende Beschreibungen handelt, die schließlich zum spezifisch phonetisch-phonologischen Teil der Arbeit überleiten sollen.

2.1 Definitionen

2.1.1 Sprechtempo

Unter dem Begriff „Sprechtempo“ werden in dieser Arbeit unterschiedliche Ausprägungen der Schnelligkeit gesprochener Sprache verstanden. Das Sprechtempo wird damit als messbare Geschwindigkeit aufgefasst, die in verschiedenen Einheiten (z.B. Laute, Silben, Wörter oder Morpheme pro Sekunde; ab hier „/s“) ausgedrückt werden kann. Des Weiteren wird diese messbare Geschwindigkeit als vergleichbar dargestellt, so dass eine Anzahl von Ereignissen gezählt und damit die Schnelligkeit ihrer Abfolge mit der in einem weiteren Zeitraum verglichen werden kann. Neben dem Konzept der produzierten sprachlichen Elemente pro Zeiteinheit existiert das der wahrgenommenen sprachlichen Elemente pro Zeiteinheit, die wahrgenommene Sprechgeschwindigkeit (s. Kap. 5). Grundsätzlich kommen für die Messung der Geschwindigkeit sowohl unterschiedliche zählbare Ereignisse als auch unterschiedliche zeitliche Begrenzungen in Frage. Zunächst sollen zwei zeitliche Rahmen gegenübergestellt werden: Sprech- und Artikulationszeit.

2.1.1.1 Sprech- und Artikulationsgeschwindigkeit

Die Sprechzeit (SZ) schließt alle sprachlichen Elemente innerhalb einer Äußerung ein. Dazu zählen das akustische Sprachmaterial inklusive der darin enthaltenen Pausen, auch Hästitionsphänomene wie Fülllaute und Dehnungen, Schnalz- und Sauglaute und andere teilweise von der Artikulation unabhängige Geräusche. Derartige Laute können unterschiedlich auf den Inhalt der Äußerung bezogen sein. So weist der Schmatzlaut, den einige Sprecher nach Pausen produzieren, keine inhaltliche Bedeutung auf, wohingegen z.B. eine verzögernde Dehnung der ersten Silbe in „könnte“ inhaltlich den Grad einer Wahrscheinlichkeit modifizieren kann. Wie häufig diese Mittel eingesetzt werden, hat einen direkten Einfluss auf die SZ. In Abbildung 2.1 ist ein Abschnitt der Stille markiert, der unerwarteterweise zwischen Subjekt und Verb produziert wird und Spannung erzeugt. Als rhetorischer Zweck dieser Pause kann eine Hervorhebung des folgenden Satzglieds vermutet werden.

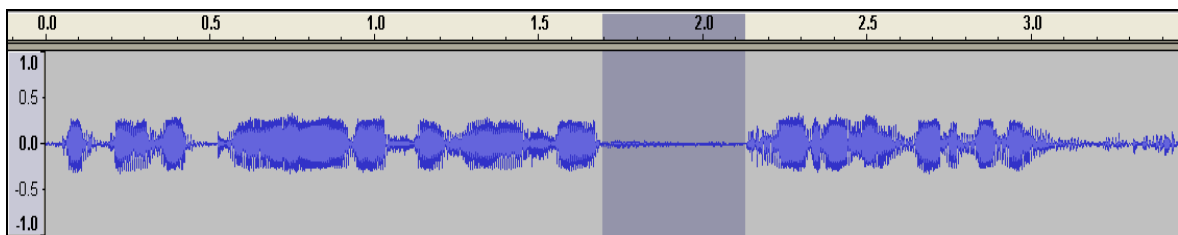


Abbildung 2.1: SZ inkl. markiertem Abschnitt der Stille zwischen Subjekt und Verb in ep. „A Orey Financial [0,444 s Pause] aguarda luz verde do Banco de Portugal“ (de. „Orey Financial [Pause] wartet auf grünes Licht der portugiesischen Zentralbank“), s. 4.4.7.4. zu diesem Beispiel.

Von der SZ unbedingt zu unterscheiden ist die Artikulationszeit (AZ). Innerhalb der AZ findet immer Artikulation statt, d.h. stille Pausen werden nicht als Bestandteil der AZ gerechnet. Dabei muss festgelegt werden, wie „gefüllte Pausen“ (s. 2.4 zu diesem Begriff) gewertet werden, in denen auch Laute produziert werden können, die nicht zur Enkodierung des Inhaltes einer Nachricht eingesetzt werden. Wenn gemessen werden soll, wie viel Zeit von einem Sprecher in einer Situation benötigt wird, um einen bestimmten Inhalt artikulatorisch umzusetzen, liegt es nahe, die Zeit, die z.B. für die kognitive Vorbereitung (Wortwahl) verwendet wird, und damit Füllwörter und Fülllaute, unberücksichtigt zu lassen. Psychologisch betrachtet zählt die für Hästitionen verwendete Zeit zu der Zeit, die für die Enkodierung (als Teil der Sprachproduktion) verwendet wird. In welchen Fällen produzierte Laute psychologisch relevant sind bzw. rhetorische oder diskursstrukturierende Funktion aufweisen, kann der Beobachter jedoch nie mit

endgültiger Sicherheit feststellen, was auf die umstrittene Wertung dieser Häsitationen hinweist.

Das in der SZ gemessene Tempo wird als Sprechgeschwindigkeit (SG), das in der AZ gemessene Tempo als Artikulationsgeschwindigkeit (AG) bezeichnet.¹ Dabei bezieht sich der Begriff „Artikulation“ auf die Aktivitäten der Artikulationsorgane, wohingegen der Begriff „Sprechen“ einen stärkeren Bezug zum Inhaltlichen hat. Unter der SG soll hier die Schnelligkeit des Kommunizierens propositionalen Gehaltes verstanden werden. Im ersten Teil dieser Arbeit wird das Sprechtempo in erster Linie unter artikulatorischen Aspekten betrachtet, wobei auch inhaltliche Elemente berücksichtigt werden.

Weiterhin ist der Begriff der Schnelligkeit auch auf der Diskursebene definierbar. Ein denkbarer Untersuchungsgegenstand ist hier die Länge der Redebeiträge und die Häufigkeit von Sprecherwechseln, was in dieser Arbeit nicht berücksichtigt wird.

2.1.1.2 Zählbare sprachliche Elemente

Es stellt sich die Frage, welche Elemente sinnvollerweise in einem gegebenen Zeitraum gezählt werden sollen. Prinzipiell kommen sprachliche Elemente unterschiedlichster Größen in Betracht, in den meisten Untersuchungen werden Elemente zwischen der segmentalen und der phrasalen Ebene gezählt. Laute scheinen als zu zählende Elemente angemessen, um die Schnelligkeit der Artikulation darzustellen, da mit ihnen die Abfolge kleinster phonetischer Elemente betrachtet wird,² bedürfen jedoch einer exakten Definition.

Als weiteres sprachliches Element zur Messung des Sprechtempos dient das Wort, das sehr unterschiedliche Definitionen aufweist (orthografisch, morphologisch, semantisch, syntaktisch, phonologisch, psychologisch) und im Vergleich zum Laut allgemein als stärker inhaltsbezogen bezeichnet werden kann. Daher eignet es sich, um auszudrücken, wie viel semantischer Gehalt in einem Zeitraum geäußert wird, also als Element zur Messung der SG, nicht der AG.

Zwischen diesen beiden Größen befindet sich die Silbe, die als phonologisches Element auch geeignet ist, die AG zu beschreiben. Eine mögliche phonetische Definition bleibt umstritten.

¹ In einigen Untersuchungen werden hiervon abweichende Begriffe verwendet, z.B. bezeichnet Meinhold (1967: 108) die SG als „Verstehensgeschwindigkeit“ oder „Apperzeptionsgeschwindigkeit“.

² Dabei soll nicht unberücksichtigt bleiben, dass diese Einheiten weder diskret noch invariant sind und außerdem als separat „gespeicherte“ Elemente bezweifelt werden, vgl. u.a. Ladefoged (2005: 186).

Ein stark inhaltsbezogenes sprachliches Element ist das Morphem. Die Einheit Morpheme/s gibt an, wie viel semantische und grammatische Information in einem zeitlichen Abschnitt ausgedrückt wird, wobei sich der Umfang der grammatischen Information der Intention des einzelnen Sprechers entzieht. Diese Einheit kann im weiteren Sinne als Informationsdichte³ bezeichnet werden und gibt Auskunft darüber, wie ökonomisch Inhalte mit sprachlichen Elementen ausgedrückt werden. Meinhold (1967: 107) definiert die Informationsdichte als „semantisches Gewicht“ phonetischen Materials, insbesondere in Bezug auf die Umgebung von Pausen, für die diese Informationsdichte in der Wahrnehmung eine besondere Rolle spielen. Die mittlere Informationsdichte (hier: Morphem/s) hängt vom morphologischen Typ der untersuchten Sprache ab, der wiederum vom phonologischen Typ beeinflusst sein kann (vgl. Fenk-Oczlon/Fenk, 2006; s.u.).

Im Sprachvergleich ergibt sich das Problem, dass alle genannten Größen sehr unterschiedlich ausfallen können (s. Bsp. (2.1) weiter unten zum Vergleich eines deutschen Wortes mit einem spanischen mit gleicher Bedeutung). Wörter können in agglutinierenden Sprachen wie dem Finnischen oder Türkischen verhältnismäßig lang werden, wohingegen in isolierenden Sprachen wie dem Chinesischen die meisten Einsilber als potentiell eigenständige Wörter benutzt werden. Auch im Deutschen lassen sich auffallend lange Wörter finden. Meist handelt es sich dabei jedoch um seltene Komposita, wie bei dem konstruierten „Donaudampfschiffahrtskapitäns-...“, das im tatsächlichen Sprachgebrauch praktisch nicht vorkommt. Nicht nur für agglutinierende Sprachen ist die Definition des Wortes problematisch, wie z.B. Bauer (2000: 252) anhand von Beispielen aus dem Finnischen aufzeigt, bei denen orthografische und phonologische Wörter differieren. Vergleicht man zwei Sprachen, von denen eine tendenziell kürzere, die andere tendenziell längere Wörter verwendet, kann die Einheit Wörter/s ein verzerrtes Bild der SG wiedergeben. Sinnvoller erscheint hier die Einheit Wörter/Phrase (bzw. pro Satz oder artikulatorischer Phrase⁴), die jedoch bei natürlichem Sprachmaterial nur schwer kontrollierbar und damit kaum vergleichbar zu messen ist.

Auch die Silbe weist im Sprachvergleich sehr unterschiedliche durchschnittliche Größen auf (s.u.). Sprachen können stark ausgeprägte Tendenzen zu einfachen, aber auch zu

³ In Anlehnung an den physikalischen Dichtebegriff ($\rho = \text{Masse/Volumen}$) wäre die Informationsdichte im sprachlichen Kontext als Anzahl der Morpheme innerhalb des Sprachmaterials zu definieren. Etwas vereinfacht wird darunter in Anlehnung an den Sprechtempobegriff hier die Anzahl der in einer Äußerung enthaltenen Morpheme in einem Zeitraum verstanden. Über den Informationsgehalt einzelner Morpheme soll damit nichts ausgesagt werden.

⁴ Unter dem Begriff „artikulatorische Phrase“ wird in dieser Arbeit der Abschnitt zwischen zwei Pausen verstanden.

komplexen Silbenstrukturen aufweisen, wodurch im Sprachvergleich die Einheit Silben/s eher die durchschnittliche Silbenlänge, nicht aber notwendigerweise die Schnelligkeit bei der Artikulation darstellt.

Schließlich können auch auf der lautlichen Ebene quantitative Unterschiede zwischen Sprachen existieren. Vokale z.B. können, abgesehen von der inhärenten phonetischen Länge, in einigen Sprachen phonologisch generell kurz sein, in anderen Sprachen einen Längenkontrast aufweisen, der wiederum unterschiedlich stark ausgeprägt sein kann, was die phonetische Dauer anbelangt. Auch Kompressionen bei ansteigender Zahl der Laute in der Silbe verursachen Unterschiede in der Lautdauer. Für die phonetische Untersuchung der AG ist es wünschenswert, Effekte, die von morphologischen oder phonologischen Gegebenheiten einer Sprache stark beeinflusst sind, möglichst kontrolliert zu halten.

(2.1)⁵

	DE	ES	Unterschied (quantitativ)
	<Strumpf>	<calcetín>	
Laute ⁶	7: [ʃtʁʊmpf]	8: [kaɫθe'tin]	Es. = De. · 1,14
Silben	1: /ʃtʁʊmpf/	3: /kaɫ.θe.'tin/	Es. = De. · 3
Wörter	1: Strumpf	1: calcetín	Es. = De. · 1
Morpheme	1: {Strumpf}	2: {Strickarbeit} {DIM}	Es. = De. · 2

In (2.1) liegt in beiden Wörtern die gleiche Bedeutung vor,⁷ allerdings umfasst das deutsche Wort ein, das Spanische zwei Morpheme, also doppelt so viele. Sieben Laute im Deutschen stehen acht Lauten im Spanischen gegenüber, was ein geringer Unterschied ist, wenn man berücksichtigt, dass die Silbenzählung die dreifache Anzahl im Spanischen ergibt. In diesem Beispiel verwendet das Spanische also etwas mehr Laute, die doppelte Anzahl von Morphemen und drei Mal so viele Silben wie das Deutsche. Der Unterschied zwischen den Sprachen bezüglich der gemessenen AGs fällt in Abhängigkeit der Verwendung verschiedener Einheiten sehr unterschiedlich aus: Angenommen, ein deutscher und ein spanischer Sprecher sprechen mit einer identischen AG von zwölf Lauten/s, so ergibt sich basierend auf dem Beispiel für den Sprecher des Deutschen eine in

⁵ Nummerierungen entsprechen bei Zitaten prinzipiell nicht der Nummerierung des Originals, sondern der der vorliegenden Arbeit.

⁶ Für [pf] (teilweise auch als [pʰf] dargestellt) werden zwei Segmente, genauer: Artikulationsbewegungen, gezählt, nämlich ein Verschluss und eine Reibung.

⁷ De. „Strumpf“ und es. „calcetín“ sind nicht vollkommen gleichbedeutend: „Calcetín“ bezieht sich auf (kurze) Socken, wohingegen der deutsche „Strumpf“ lang oder kurz sein kann, weshalb „Strumpf“ oft mit „media“ und „Socke“ mit „calcetín“ übersetzt wird: vgl. jedoch den Eintrag im *Diccionario de la lengua española* der RAE zu „calcetín“. Hier wurden zur anschaulichen Verdeutlichung der strukturellen Unterschiede zwischen Spanisch und Deutsch diese inhaltlich durchaus vergleichbaren Begriffe gewählt.

Silben/s gemessene AG von 1,7 Silben/s, für den Sprecher des Spanischen ergibt sich eine mehr als doppelt so hohe AG von 4,5 Silben/s. Die Morphem- und Wortgeschwindigkeiten liegen dann im Deutschen beide bei 1,7 und im Spanischen bei 3,0 bzw. 1,5. Selbstverständlich handelt es sich hierbei nicht um Beispiele, die ohne Vorbehalte als repräsentativ für die beiden Sprachen angenommen werden dürfen, jedoch sind die Tendenzen für komplexe Silben im Deutschen und einfache Silben im Spanischen nicht von der Hand zu weisen. Anhand dieses Beispiels ist erkennbar, wie problematisch es ist, eine der beiden Sprachen generell als schnell oder langsam zu bezeichnen. Weitere Beispiele sollen verdeutlichen, dass die Berücksichtigung anderer sprachlicher Elemente ebenfalls Komplikationen im Sprachvergleich verursachen kann.

Karlgren (1961: 672) empfiehlt: „Speed must reasonably be defined as a relation between *time* and distance or volume or mass or something corresponding“ (Hervorhebung im Original). Die von den Artikulatoren zurückgelegte Distanz zwischen Zielpositionen im artikulatorischen Ablauf kann gemessen werden, erfordert aber einen Aufwand, der den Umfang des untersuchten Sprachmaterials einschränkt und die Beschaffenheit des Materials beeinflusst. Als „mass unit“ schlägt Karlgren Morpheme vor,⁸ definiert die SG also nicht über die Bewegungen der Artikulationsorgane, sondern über den Inhalt des Gesagten. Aus Bsp. (2.2) geht anhand einer synchron orientierten Morphemanalyse hervor, dass Sprachen sehr unterschiedliche Arten von Morphemen aufweisen und diese unterschiedliche Häufigkeiten aufweisen, was den Sprachvergleich wiederum erschwert. Im Chinesischen z.B. sind grammatische Morpheme selten, da die entsprechende Information in Abhängigkeit vom Kontext erschlossen wird, während sie z.B. in den romanischen Sprachen formal ausgedrückt wird. Die vergleichsweise niedrige Anzahl von grammatischen Morphemen scheint im Chinesischen durch lexikalische Redundanz ausgeglichen zu werden. Jede Silbe dient potentiell als lexikalisches Morphem, umgekehrt wird in der Regel jedes Morphem durch eine Silbe ausgedrückt.

⁸ Karlgren (1961) definiert den Begriff „Morphem“ nicht weiter.

(2.2)

	CN ⁹	ES	Unterschied (quantitativ)
	< 教育 >	< educación >	
Laute ¹⁰	7: ['tɕjaɯjy]	9: [eðuka'θjon]	Es. = Cn. · 1,3
Silben	2: /'tɕjaɯ.jy/	4: /e.ðu.ka.'θjon/	Es. = Cn. · 2
Morpheme	2: {lehren} {erziehen}	2: {lehren/erziehen} {DER ^{V→N} }	Es. = Cn.
Bedeutung	„erziehen / Erziehung“	„Erziehung“	

Während die beiden Morpheme im Chinesischen in nur zwei Silben ausgedrückt werden, verwendet das Spanische hierfür vier Silben, was eine niedrigere Informationsdichte zur Folge hat. Gleichzeitig wird im Chinesischen eine Bedeutung durch zwei annähernd bedeutungsgleiche Morpheme ausgedrückt (vgl. Bsp. (2.3) für ein Wort im Chinesischen, das aus einer Silbe und einem Morphem besteht). Lexikalische Morpheme können hier relativ frei zu Wörtern kombiniert werden. Derivationsmorpheme gibt es kaum, die häufigste Art grammatischer Wortklassenwechsel ist die Konversion ohne formelle Mittel. Im Spanischen wird die Wortart morphologisch markiert. Nichtvorhandene grammatische Markierungen im Chinesischen bewirken andererseits eine relativ niedrige Morphemdichte.

(2.3)

	CN	ES	Unterschied (quantitativ)
	唱	cantas	
Laute	5: [tʂ ^h aŋ]	6: ['kaŋtas]	Es. = Cn. · 1,2
Silben	1: /tʂ ^h aŋ/	2: /'kan.tas/	Es. = Cn. · 2
Morpheme	1: {sing}	3: {singen} {Präs.Ind.} {2.P.Sg.}	Es. = Cn. · 3
Bedeutung	<i>kontextsensitiv</i>	„du singst“	Es. ≈ Cn.

Hier weist das Spanische durchschnittlich 1,5 Morpheme/Silbe bzw. 3 Morpheme/Wort auf, das Chinesische jeweils 1,0. Aussagen über die geschätzte Morphemgeschwindigkeit können damit vorerst nicht gemacht werden.

In den meisten Arbeiten zum Sprechtempo wird davon ausgegangen, dass *eine* Einheit angemessen ist, das Sprechtempo darzustellen (z.B. bei Abercrombie, 1967: 96, die Silbengeschwindigkeit; Osser/Peng, 1964 ziehen die Lautgeschwindigkeit vor). In dieser Arbeit wird die Ansicht vertreten, dass zum Zweck des Sprachvergleichs viele unterschiedliche Messgrößen betrachtet und als unterschiedliche Parameter des

⁹ Die phonologischen Töne des Chinesischen werden in dieser Arbeit nur dann transkribiert, wenn sie zum Verständnis notwendig sind, ansonsten wird zur Optimierung der optischen Gestalt darauf verzichtet.

¹⁰ Die Aussprache mit silbeninitialem [j] in der zweiten Silbe wird von manchen Sprechern bevorzugt, was hier in der Lautzählung berücksichtigt wird.

Sprechtempos miteinander verglichen werden müssen. Es wird auch davon ausgegangen, dass keine linearen Beziehungen zwischen der Geschwindigkeit, mit der artikuliert wird und der Geschwindigkeit, mit der kommuniziert wird, vorliegen müssen. Für die genannten Einheiten muss eine genaue Definition der gezählten Größe geliefert werden, die für alle behandelten Sprachen gültig ist (s. 4.4.2).

Als weitere zählbare sprachliche Elemente, die in dieser Arbeit nicht berücksichtigt werden können, sind z.B. Scollons Schläge/min zu erwähnen.¹¹ Schläge (*Beats*) entsprechen akzentuierten Silben und beeinflussen die wahrgenommene SG. Scollon unterscheidet zwischen „rate“ (SG) und „density“ (der Anzahl von akzentuierten Silben) und betont, dass diese Größen voneinander unabhängige rhythmische Eigenschaften darstellen. Uhmann (1992) zeigt, wie verschiedene Kombinationen aus SG („Dichte 1“) und Häufigkeit betonter Silben („Dichte 2“) u.a. Parenthesen und Relevanz kontextualisieren. Da Grad und Häufigkeit von akzentuierten Silben vielen nur schwer kontrollierbaren Einflüssen obliegen und Akzente außerdem unterschiedlich phonetisch realisiert werden können, wird hier auf die Untersuchung dieser suprasegmentalen Elemente verzichtet. Eine grundsätzliche Unterscheidung zwischen SG (Silben/s ohne weitere Unterteilung in SG und AG und ohne Erwägung anderer Elemente als Silben) und Dichte (Verhältnis zwischen betonten und unbetonten Silben) wird von Auer et al. (1999) in einer an Kontextualisierung in Gesprächen orientierten Analyse von Rhythmus und Tempo vorgenommen.¹²

Zur Erfassung des Sprechtempos wurden auch Mora/s gezählt, z.B. in Ohno et al. (1996), was für den Vergleich der hier behandelten Sprachen jedoch nicht weiter relevant ist, da die Definition der Größe außerdem von den Silbengewichtsverhältnissen innerhalb von Einzelsprachen abhängig ist.

Eine Auszählung von artikulatorischen Phrasen pro Äußerung wird nicht angestrebt; dafür wird in Form einer Analyse der Pausenstrukturen angegeben, wie oft Pausen produziert werden und wie lange, in absoluter Zeit und Silben gemessen, artikulatorische Phrasen sind.

Eine denkbare Definition des Sprechtempos im Gespräch ist die Geschwindigkeit von *Turn*konstruktionen. Es soll an dieser Stelle lediglich angemerkt werden, dass aufgrund von unterschiedlichen Diskursregeln für Gespräche unter Japanern eine langsamere Abfolge von Gesprächsbeiträgen vermutet wird, da Redebeiträge tendenziell nicht

¹¹ S. Uhman (1997: 185 ff.).

¹² Weitere Literatur zum Konzept der Dichte ist bei Uhmann (1992; 1997) und Selting (1995) zu finden; zur Diskussion über die Rhythmusklassifikation s. Kap. 3.

unterbrochen werden, wohingegen in vielen westlichen Gesellschaften häufige Unterbrechungen und simultane *Turns* eine schnelle Abfolge von Gesprächsbeiträgen ermöglichen können (vgl. Nakane, 2007 und Literaturangaben).

Der Begriff „Sprechtempo“ wird in der vorliegenden Arbeit als allgemeiner Begriff immer dann verwendet, wenn nicht zwischen einzelnen Parametern (z.B. Laut- und Silbengeschwindigkeit) bzw. zwischen AG und SG unterschieden werden muss.

2.1.2 Zeit

Wenn gemessen wird, wie viele sprachliche Elemente innerhalb eines Zeitraumes produziert werden, muss zuvor auf unterschiedliche Definitionen des Begriffs „Zeit“ eingegangen werden. In den Naturwissenschaften wurde das von Newton geprägte Konzept der Zeit als absolute Größe, die unabhängig von anderen Größen messbar ist, aufgrund der Erkenntnisse der Relativitätstheorie verworfen. Zeit gilt als durch die relative Bewegung verschiedener Beobachter zueinander, aber auch durch Massen beeinflussbar.¹³ Auch das menschliche Empfinden von Zeit ist relativ. Verschiedene Sinnesorgane weisen unterschiedliche Schwellen bezüglich der Wahrnehmung von Reizen als diskrete Einheiten auf: Für die auditive Wahrnehmung liegen diese niedriger als für die optische Wahrnehmung.

Neben der Auffassung von Zeit als messtechnisch erfassbare Größe kann diese auch als subjektiv wahrgenommene Größe verstanden werden – die „erlebte Zeit“ muss nicht mit der gemessenen Zeit übereinstimmen, sondern kann u.a. von der Anzahl und der Intensität kognitiver, aber auch emotionaler Prozesse beschleunigt bzw. verlangsamt werden. Folglich kann die wahrgenommene SG in Abhängigkeit des Inhaltes einer Aussage variieren. Es gibt auch Hinweise darauf, dass von der inneren Verfassung unabhängige situative Umstände die erlebte Zeit beeinflussen. DeLong (1981) fasst Experimente zusammen, die darauf hinweisen, dass Versuchsteilnehmer, die sich in verkleinerte Umgebungen hineinversetzen, ein dem Maßstab entsprechendes Zeitgefühl entwickeln. Dabei werden insgesamt 449 Versuchspersonen mit verkleinerten Räumen (Wohnzimmer, Wartezimmer, usw.) konfrontiert, wobei sie spezielle Masken tragen, die die Wahrnehmung der Umgebung außerhalb dieser Schauorte verhindern. Sie erhalten die Anweisung, sich vorzustellen, genau 30 Minuten in den präsentierten Räumlichkeiten zu

¹³ Dabei wird unter „Beobachter“ kein denkendes oder empfindendes Wesen verstanden, vgl. Dowden (2001).

verbringen und dann ein Signal zu geben. Die Ergebnisse zeigen, dass die durchschnittliche gemessene Zeit dabei einen dem Maßstab der verkleinerten Räume entsprechenden Teil der 30 Minuten beträgt (getestete Maßstäbe: 1:24, 1:12, 1:6 und 1:1), wobei dieser Effekt nur bis zu einem Maßstab von 1:12 beobachtet werden kann. Die Erkenntnis, dass subjektiv erlebte Zeit relativ ist, hat Auswirkungen auf das Verständnis von Tempo, das ebenfalls als subjektive Größe bezeichnet werden muss. Stern (1900: 115) versteht unter dem Begriff Tempo „nicht nur ein objektives Melodie-, sondern stets auch ein subjektives Auffassung-Tempo“. Für das Sprechen betont Stern die Un-Messbarkeit des Tempos: „das Sprechen ist zu kompliziert und allzu vielen Bedingungen unterworfen, als dass es ein allgemeines Tempomaass abgeben könnte“ (ebd., S. 116 f.). Er lässt dagegen Versuchspersonen einen dreiteiligen Rhythmus (‘ - -) klopfen, um das psychische Tempo¹⁴ zu untersuchen. In ähnlichen Untersuchungen zum präferierten Tempo, in denen Versuchsteilnehmer angewiesen werden, in einem selbst gewählten, angenehmen Tempo mit den Fingern oder Gliedmaßen zu tippen, wird ein mittleres Intervall von etwa 600 ms festgestellt:

„Although there is a great deal of intersubject variation in these personal tempos, a mean value across subjects also tends toward 600 ms. For example, Semjen, Vorberg, and Schulze (1998) obtained preferred tempo rates from 428.8 ms to 725 ms, with a mean around 565.3 ms (see also Fraisse 1982). Interestingly, Drake, Jones, and Baruch found that preferred tempo seems to change with age with young children having a faster spontaneous tempo, around 400 ms for 4—6-year-olds (2000, p. 266). If personal tempo is kinematically grounded, it makes sense that this change should occur as a reflection in the changing physical dynamics (principally size) of our bodies during different stages of development.“ (London 2004: 31)

London weist auch auf die Bedeutung eines 600-ms-Intervalls für andere Untersuchungsgegenstände hin, z.B. bei der Wahrnehmung von rhythmischen Impulsen, die bei Abständen von ungefähr 600 ms am besten wahrgenommen werden. Lenneberg (1986: 139 ff.) dagegen nimmt einen Rahmen von 120 ms als grundlegendes Intervall für (neuro-)psychologische Prozesse und präferiertes Tempo an, wobei die Ausführungen nicht auf überzeugenden Beweisen beruhen. Ungeachtet der Größe zeitlicher Intervalle, auf denen physische und/oder psychische Ereignisse basieren mögen, muss eine Unterscheidung zwischen *absoluter* und *subjektiver* Zeit getroffen werden. Selbst wenn, zumindest innerhalb bestimmter Kulturen¹⁵, eine gewissermaßen einheitliche Präferenz

¹⁴ Vgl. Fraisse (1982: 153), der den Begriff „spontanes Tempo“ verwendet und auf die Begriffe „persönliches Tempo“ bei Frischeisen-Köhler (1933) und „mentales Tempo“ bei Mishima (1951/52) hinweist.

¹⁵ Zur Definition des Begriffs „Kultur“ s. Kap. 8.

bezüglich der Wahrnehmung absoluter Zeit besteht, werden unterschiedliche Tempi in Abhängigkeit von der gelenkten Aufmerksamkeit ausgelöst, wie London (2004: 33, auf Meyer/Palmer, 2001 verweisend) in Bezug auf die subjektive Zeit unterstreicht. So wurde festgestellt, dass ein Musikstück schneller gespielt wird, wenn die Spieler sich auf halbe Noten konzentrieren, wohingegen das Tempo verringert wird, wenn Achtelnoten im Zentrum der Aufmerksamkeit stehen – obwohl es sich um dasselbe Stück bzw. denselben Abschnitt in einem Stück handelt.

Für die subjektive Gegenwart allerdings (die als das „Jetzt“ empfunden wird) wird davon ausgegangen, dass diese nicht vom kulturellen Hintergrund beeinträchtigt wird (van der Meer 2006: 83).

2.1.3 Definition von „schnell“ und „langsam“

Zwei Ansätze kommen bei der Beurteilung des Tempos gesprochener Sprache in Betracht. Zum einen können Prädikate wie „mittleres Tempo“, „erhöhtes Tempo“, „deutlich erhöhtes Tempo“ etc. nach Kardinalwerten vergeben werden. Dafür müssen konkrete Grenzen festgelegt sein, z.B. eine AG von 14 Lauten/s als Grenze von „mittlerem“ zu „erhöhtem“ Tempo. Bei der Begründung dieser Grenzen kann wiederum unterschiedlich vorgegangen werden. Einerseits könnte man davon ausgehen, dass das Sprechtempo eine normalverteilte Größe ist und bestimmte Punkte auf der Gaußschen Kurve als Grenze zwischen den Kategorien dienen, z.B. die Grenzen des 68,27%-Intervalls zur Abgrenzung „mittleren“ Tempos von „hohem“ bzw. „niedrigem“ Tempo. Dabei ergibt sich zum einen das Problem, in wie viele Kategorien eine Unterteilung sinnvoll ist. Zum anderen müssen die Grenzen auf einer Stichprobe beruhen, die ausreichend groß ist um überhaupt Rückschlüsse auf die Grundgesamtheit zuzulassen. Eine derartige Datenerhebung erscheint weder sinnvoll noch umsetzbar, wenn die Benutzer von Lautsprache ganz allgemein als Grundgesamtheit angenommen werden. Da situative Aspekte einen Einfluss auf das Sprechtempo haben, kann nicht davon ausgegangen werden, dass anhand von kardinalen Werten allgemeingültige Kategorien abgegrenzt werden können – in einer festlichen Ansprache mag als hohes Tempo gelten, was für Nachrichtensendungen ein untypisch langsames Tempo darstellt. Geht man andererseits nicht von der Artikulation als Basis für eine Einteilung in Kategorien aus, kann die Wahrnehmung, also die Beurteilung des Tempos durch Hörer, als Kriterium herangezogen werden. Es kann aber nicht ausgeschlossen werden, dass Sprecher unterschiedlicher Sprachen unterschiedliche Kategorien

aufstellen (also Sprecher der Sprache A bereits ab 14 Lauten/s, Sprecher der Sprache B erst ab 16 Lauten/s das Prädikat „erhöhtes Tempo“ vergeben).

Bei beiden Methoden muss auf Grenzen menschlicher Fähigkeiten im motorischen und perceptiven Bereich eingegangen werden. Ab einer gewissen Lautgeschwindigkeit können Lautketten nicht mehr deutlich artikuliert werden, d.h. das Erreichen aller *Targets* wird praktisch unmöglich und die Verständlichkeit ist damit nicht mehr gegeben. Calvert (1992: 178) gibt eine obere artikulatorische Grenze von acht Silben/s an, Goldman Eisler (1968: 24) erwähnt Werte aus Miller (1951), die zwischen 6,7 Silben/s für Artikulationen im hinteren Mundraum und 8,2 Silben/s für Artikulationen mit der Vorderzunge („tat, tat, tat“) liegen.¹⁶ In einem Überblick in Kowal (1991: 99) werden ähnliche Werte zitiert: Bei Hudgins/Stetson (1937) 8,2 Silben/s (Vorlesen), bei Lenneberg (1986) 8,33 Silben/s (Spontansprache) und bei Grosjean/Deschamps (1975) sogar ein Wert von 9,8 Silben/s als höchste AG, sowie bei Kaiser (1939) 10 Silben/s (Vorlesen). In einer weiteren Arbeit von Rebollo Couto (1998) wird für das Spanische die Kategorie „sehr schnell“ erst ab einem Tempo von zehn Silben/s aufgestellt. Auch in Janse (2003a) werden zehn Silben/s erreicht. Sie berichtet außerdem, dass Sprecher bei einer Beschleunigung keine Verdoppelung der AG erreichen. Soll die Schnelligkeit der Artikulationsorgane beurteilt werden (wie viele Bewegungen/s werden ausgeführt?), so sind die erwähnten Werte aus oben genannten Gründen nur eingeschränkt brauchbar. Der Wert von zehn bis 15 Lauten/s als obere Grenze der AG in Laver (1994) erscheint eher unrealistisch.

Was die Wahrnehmung anbelangt, wurden folgende Grenzwerte angegeben: bei Calvert (1992: 178) 30 Laute/s, bei Liberman et al. (1967: 432, auf Orr/Friedman/Williams, 1965 verweisend) 400 Wörter/min, was etwa 30 Lauten/s entspricht.¹⁷ Janse (2003b) berichtet von Wahrnehmungstests, in denen Aufnahmen, die mit bis zu dreifacher Geschwindigkeit abgespielt werden, noch als (zumindest teilweise) verständlich galten. Die Verständlichkeit beschleunigter Sprache scheint davon abzuhängen, ob es sich um natürliches, mit erhöhter SG gesprochenes Material oder künstlich komprimiertes Material handelt (s. de Haan, 1977, nach dem komprimierte Aufnahmen bei höherer Geschwindigkeit besser wahrnehmbar sind als natürlich schneller produzierte; auch Chang, 2005).

¹⁶ Die Untersuchung von C₁VC₁-Silben ist hierbei nicht ganz einleuchtend, wenn man die Studie Stetsons (1951) berücksichtigt, aus der hervorgeht, dass derartige Silben ab einer bestimmten AG zu CV-Silben vereinfacht werden („pup, pup, pup“ > „pu, pu, pu“; „top, top, top“ > „to, to, to“).

¹⁷ Gleichzeitig zitieren sie Miller/Taylor (1948) und erklären, dass diskrete Lautereignisse bei dieser Geschwindigkeit nicht mehr getrennt voneinander wahrnehmbar seien. Damit gehen die Autoren davon aus, dass die Sprachwahrnehmung nicht auf der Verarbeitung segmentierbarer Laute beruht.

Ein anderer Ansatz geht nicht von Grenzen aus, die mit ordinalen Werten angegeben werden, sondern setzt gemessene Werte zueinander in Bezug. „Schnell“ und „langsam“ sind dabei relativ und nur im Kontext von Vergleichswerten zu verstehen: In Sprache A ist eine bestimmte Ausprägung des Sprechtempos, z.B. die AG, als schnell im Vergleich zu Sprache B zu bezeichnen, wenn eine höhere Anzahl von gemessenen Größen pro Zeiteinheit vorliegt, aber langsam im Vergleich zu Sprache C, in der eine noch höhere Anzahl gemessen wird; keine der Sprachen ist an sich schnell oder langsam. Diese Herangehensweise billigt allen Sprachen eigene Maßstäbe zu, die unterschiedlich ausdifferenziert sein können. Mit diesem Ansatz wird impliziert, dass bei Verwendung unterschiedlicher Messgrößen verschiedene Ergebnisse zu erwarten sind, z.B. ist Sprache A schneller als Sprache B, wenn Laute/s gemessen werden, aber langsamer, wenn Silben/s gezählt werden. Für Sprache A gilt in diesem Fall eine relativ hohe Lautgeschwindigkeit bei gleichzeitig relativ niedriger Silbengeschwindigkeit.

Da es sich bei dieser Arbeit um einen Vergleich zwischen Sprachen und Kulturen handelt, wird vom zweiten, dem relativierenden Ansatz ausgegangen.¹⁸ Laver (1994: 541) empfiehlt, von den Kategorien „**medium, fast and slow rates**“ auszugehen und zwei Arten der Veränderung des Sprechtempos anzunehmen: „**accelerating and decelerating rates**“ (Hervorhebungen im Original). Er gibt keine exakten Grenzen für diese Kategorien an und geht damit von relativen Prädikaten aus, nennt aber als „Faustregel“ einen Bereich von 160 bis 240 Wörtern/min als Normalbereich. Teilweise stehen Angaben zur normalen durchschnittlichen SG einer Sprache im Widerspruch zueinander. Calvert (1992: 178) geht von einer mittleren Silbendauer von 0,18 s für amerikanisches Englisch aus, was einer Geschwindigkeit von 5,0 bis 5,5 Silben/s entspricht. Pickett (1980: 9) dagegen setzt (ebenso für amerikanisches Englisch) eine durchschnittliche Silbenlänge von 0,3 s an und bezeichnet damit ein Tempo von 3,33 Silben/s als normal. Derartige Diskrepanzen können möglicherweise damit erklärt werden, dass unterschiedliche Sprechsituationen untersucht wurden.

Fenk-Oczlon/Fenk (2006) greifen die Menzerathschen Gesetze auf (s. Menzerath, 1954: 96 ff.¹⁹) und weisen darauf hin, dass zwischen verschiedenen Maßeinheiten der AG Korre-

¹⁸ Arbeiten, die ordinale Werte zur Abgrenzung von Sprechtempi verwenden, sind u.a. von Essen (1949) für Deutsch, den Os (1988) für Italienisch und Niederländisch, Gros et al. (1999) für Slowenisch, Rebollo-Couto (1998) für Spanisch, Block/Killen (1996) für australisches Englisch, Robb et al. (2004) für neuseeländisches und amerikanisches Englisch uvm.

¹⁹ Zusammengefasst sei hier die „Sparsamkeitsregel“ auf S. 101 (ebd.) widergegeben: „[J]e größer das Ganze, um so kleiner die Teile“ (hier ohne Hervorhebung des Originals).

lationen bestehen. In ihrer Untersuchung von 34 indoeuropäischen und nicht-indoeuropäischen Sprachen ergibt sich eine negative Korrelation zwischen der Anzahl von Silben pro Satz und der Anzahl von Lauten pro Silbe. Der Inhalt der Sätze ist dabei gleich, es handelt sich allerdings um Übersetzungen. Untersucht wird also, wie ein bestimmter Inhalt in unterschiedlichen Sprachen in Sprachmaterial enkodiert wird. Es ist daher zu erwarten, dass Sprachen, die eine niedrige Silbendichte aufweisen, diese im Bereich der Lautdichte, und folglich der Lautgeschwindigkeit, ausgleichen und mit Sprachen, die eine hohe Silbendichte aufweisen, vergleichbare Ergebnisse liefern. Voraussetzung hierfür wäre, dass das Sprachmaterial der Einzelsprachen vergleichbare zeitliche (z.B. Länge der Segmente) und weitere linguistische (z.B. Akzentverhältnisse, Fokussierung) Eigenschaften aufweist. Beispiele für Sprachen bei Fenk-Oczlon/Fenk mit Extremwerten sind Niederländisch mit etwas mehr als fünf Silben pro Satz²⁰, die durchschnittlich etwas weniger als drei Laute umfassen, und Japanisch mit über zehn Silben pro Satz, die durchschnittlich weniger als zwei Laute aufweisen. In der Studie liegen aber bezüglich der absoluten Lautzahl Abweichungen vor: Die Anzahl der Laute pro Satz liegt zwischen 12,5 und 19,2; Angaben zum Tempo werden nicht gemacht. Es stellt sich die Frage, was bewirkt, dass Sprachen höhere oder niedrigere Laut- und Silbendichte und -geschwindigkeiten aufweisen.

2.2 Sprachvergleich: schnelle vs. langsame Sprachen?

Was die Schnelligkeit artikulatorischer Bewegungen im segmentalen Bereich anbelangt, kann untersucht werden, ob in gewissen Sprachgemeinschaften²¹ die motorischen Fähigkeiten auf besondere Weise genutzt werden, nämlich so, dass eine außergewöhnlich hohe Lautgeschwindigkeit gepflegt wird. Sollte dies der Fall sein, kann weiterhin untersucht werden, mit welcher zugrunde liegenden Strategie eine höhere Lautgeschwindigkeit erreicht wird. Kröger (1996, 1998) erklärt zwei Möglichkeiten zur Erhöhung der AG: Die Kürzung der zeitlichen Ausdehnung einzelner Artikulationsbewegungen und die Verstärkung der Überlappung. Löfqvist (1999) geht sogar davon aus, dass der Grad der Überlappung direkt von der AG abhängt. Recasens (1999) listet einige Arbeiten auf, die stärkere Überlappung bei erhöhtem Tempo in VC-, CC- und CV-Sequenzen belegen. Die Verkürzung der zeitlichen Ausdehnung kann auf eine schnellere Artikulatorbewegung bei gleichem zurückgelegtem Weg des

²⁰ Als „clauses“ bezeichnete einfache Sätze, die einer Proposition in einer Intonationsphrase entsprechen.

²¹ Zur Definition des Begriffes „Sprachgemeinschaft“ s. Kap. 6.

Artikulationsorgans zur Zielposition zurückzuführen sein. Es kann jedoch auch eine Verkürzung des Weges stattfinden (s. Lindblom, 1963 für Vokale; Kent/Moll, 1972 für Konsonanten). Kent (1983) stellt aufgrund der Verkürzung des vom Artikulator zurückgelegten Weges bei erhöhtem Sprechtempo langsamere Artikulationsbewegungen als bei niedrigem Sprechtempo fest (s. auch Kröger, 1996; Überlegungen in Wood, 1973a und 1973b und Gay et al., 1974 zu geringerer Muskelaktivität bei Vokalen bei höherem Tempo, begleitet von *undershoot*, aber höherer Muskelaktivität bei labialen Konsonanten). Untersuchungen zu Strategien bei Beschleunigung des Sprechtempos (Kröger 1996, 1998; Byrd/Tan, 1996) ergeben, dass alle genannten Strategien nachweisbar sind, teilweise auch mehrere Strategien vom gleichen Sprecher angewandt werden. Farnetani (1990: 109 f.) fasst hierzu zusammen:

„As reported in reviews on the effects of speaking rate on the dynamic of speech movements (Ostry and Munhall, 1985; Wieneke, Janssen and Belderbos, 1987), the following articulatory changes can occur: a) a reduction in the amplitude of the movements with no changes in velocity (Kent and Moll, 1972; Wieneke et al, 1987; b) [sic] a reduction in the amplitude and an increase in velocity (Gay, 1981); c) no changes in the amplitude and an increase in velocity (Gay, Ushijima, Hirose and Cooper, 1974); d) reductions in both amplitude and velocity.“

Kann eine Sprache aufgrund häufiger Verwendung einer oder aller dieser Strategien generell eine hohe Lautgeschwindigkeit aufweisen? Eine prinzipielle Verkürzung der Artikulationsbewegungen zur jeweiligen Zielposition würde dem Prinzip der Verständlichkeit entgegenwirken, gleichzeitig dem der Sprachökonomie entsprechen. Im Falle des Spanischen fallen in diesem Kontext systematische Lautveränderungen auf, die auf eine Reduktion des artikulatorischen Aufwandes und eine Verkürzung des zurückgelegten Weges zurückgeführt werden können: / p t k / wurden zunächst intervokalisch zu / b d g / sonorisiert (Reduktion des artikulatorischen Aufwandes, die Vibration der Stimmbänder zwischen den stimmhaften Vokalen auszusetzen sowie des respiratorischen Aufwandes bei stimmlosen Lauten im Allgemeinen, vgl. Ladefoged, 1983), die sich in dieser Position zu Reibelauten entwickelten [β ð γ]²² (die Zielposition des Verschlusses wird nicht erreicht) und auf heutiger Sprachstufe zumindest in der Umgangssprache, teilweise aber auch in formeller Sprache, zum Ausfall neigen. Eine Artikulation als Reibelaut für stimmhafte Verschlusslaute ist auch für andere Sprachen bekannt. Okada (1999) beschreibt für

²² Auch Lindblom (1983) sieht für diesen Lautwandel phonologischen *undershoot* als mögliche Ursache. Eine weitere Reduktion von Reibelauten wäre die Realisierung als Approximanten. In einigen Beschreibungen werden die erwähnten Laute tatsächlich als Approximanten beschrieben, z.B. Martínez Celdrán/Fernández Planas (2007) und bei Canepari (2005) als stellungsbedingte Varianten für / b d / vor /w/.

Japanisch, dass für /g/ allgemein eine Tendenz zur Artikulation als [ɣ] in intervokalischer Position besteht und /b/ bei erhöhtem Sprechtempo als [β] artikuliert werden kann. Da hier nicht alle stimmhaften Verschlusslaute systematisch von dieser Tendenz erfasst werden (/d/ wird nicht als [ð] artikuliert) und die Veränderungen nicht eindeutig auf die AG zurückgeführt werden können, soll nicht von einem systematischen Sprachwandel, der auf Allegroformen beruht, ausgegangen werden. Sprachspezifische Unterschiede im Grad der Überlappung der Segmente, die für die oben zitierten unterschiedlichen Strategien b) - d) ausschlaggebend sein können, sind Gegenstand des Abschnitts 3.3. Das Anstreben einer hohen Lautgeschwindigkeit muss nicht zwingend als Grund lautlicher Veränderungen angesehen werden, da das erwähnte Prinzip der Sprachökonomie alleine derartige Prozesse erklärt. Schüppert et al. (2009) gehen davon aus, dass im Sprachvergleich Reduktion und Tempo positiv miteinander korrelieren und weisen einen derartigen Zusammenhang im Sprachenpaar Dänisch-Schwedisch nach. Dabei werden kanonische Silben gezählt, also für das im Dänischen einsilbig realisierte Wort „svarer“ zwei zugrunde liegende phonologische Silben. Selbstverständlich weist das Dänische dann in einem vergleichbaren Kontext ein höheres Tempo auf als das Schwedische, in dem das entsprechende „svarar“ zweisilbig ausgesprochen wird, was einen systematischen Unterschied zwischen diesen beiden Sprachen darstellt. Da im Schwedischen also inhaltlich identische Äußerungen aus umfangreicherem phonetischen Material bestehen, werden zwar gleich viele phonologische Silben produziert, allerdings hierfür ein längerer Zeitraum benötigt. Damit ist allerdings keine Aussage über die Schnelligkeit der Artikulation gemacht, was die Autoren verkennen, da sie die Anzahl der kanonischen Silben/s als AG bezeichnen. Reduktion und Geschwindigkeit werden in vielen Modellen der Sprachproduktion miteinander in Verbindung gebracht, wobei meistens davon ausgegangen wird, dass die Erhöhung des Tempos als Ursache für Reduktionen zu verstehen ist (s. Lindblom, 1983).

Nicht auszuschließen ist auch, dass in bestimmten Sprachen Präferenzen für artikulatorisch weniger aufwendige Laute vorhanden sind, die mit vergleichsweise agilen Artikulatoren gebildet werden oder eine geringe inhärente Länge aufweisen. Die Zungenspitze ist im Vergleich zum Zungenrücken oder zum Velum ein agiles Artikulationsorgan, das schnellere Bewegungen ausführen kann (s.o.). Bauer et al. (2010) messen eine Geschwindigkeit von 11,8 cm/s im labialen Bereich, 12,5 cm/s für die Zungenspitze und 8,4 cm/s für den mittleren und hinteren Bereich der Zunge. Wood (1973a) fasst Ergebnisse u.a. von Stetson (1928) und Hudgins/Stetson (1937) zusammen, nach denen sich die Artikulationsorgane folgendermaßen nach ihrer Agilität absteigend ordnen lassen: Zungenspitze, Kiefer, Zungenrücken, Lippen, Velum. Konsonanten, die mit der

Vorderzunge im vorderen Mundraum gebildet werden, können also am schnellsten artikuliert werden. Um die Ursache für unterschiedliche Lautgeschwindigkeiten zu erfahren, muss Rücksicht darauf genommen werden, inwieweit in einem sprachlichen System Artikulationsbewegungen begünstigt sind, die prinzipiell geringe zeitliche Ausdehnung in Anspruch nehmen. Daher werden in Kapitel 4 die Lautinventare und Häufigkeiten einzelner Laute der untersuchten Sprachen dargestellt.

Versteht man wie Keller (1990) perzeptive Distinktivität und Sprechtempo („speed of transmission“) als in Konkurrenz miteinander stehende Ziele,²³ können die für das Spanische typischen Lautwandelerscheinungen als Auswirkungen eines angestrebten hohen Sprechtempos interpretiert werden.²⁴ Gleichzeitig fallen systematische Veränderungen im Deutschen auf, die die Struktur der Silbe verschlechtern, die des phonologischen Wortes verbessern und durch Konsonantenepenthese den Artikulationsablauf komplexer gestalten, was auf ein niedriges angestrebtes Sprechtempo hinweisen kann (mhd. „ieman“ > frühnhd. „jemand“) oder zumindest kein erhöhtes Tempo als bevorzugt erscheinen lässt.

Abschließend wird festgehalten, dass, wenn von schnellen und langsamen Sprachen, Sprechern oder Abschnitten einer Äußerung gesprochen wird, genau beschrieben werden muss, ob damit schnelle Artikulationsbewegungen, eine hohe Informationsdichte oder etwa geringe Pausenzeit gemeint ist und ob dies relativ zu verstehen ist.

2.3 Faktoren des Sprechtempos

Da das Sprechtempo ständig zum Teil sehr starke Veränderungen widerfährt (vgl. Miller et al., 1984), die Gegenstand zahlreicher Untersuchungen sind,²⁵ ist es sinnvoll, zwischen lokalem und globalem Sprechtempo zu unterscheiden. Unter lokalem Sprechtempo wird die AG verstanden, die innerhalb kurzer Zeitfenster gemessen wird, z.B. in Silben, Wörtern und Phrasen. Obwohl ein begrenzter zeitlicher Rahmen dafür spricht, möglichst kleine sprachliche Elemente zu messen, wird in vielen Studien die Messgröße Silben/s verwendet. Lokale Veränderungen betreffen auch nur Teile von Silben, einzelne Laute.

²³ Vgl. aber Gimson (1962), der davon ausgeht, dass die Geschwindigkeit die Verständigung nicht beeinflusst, also nicht mit ihr in Konkurrenz stehen kann (nach Grunwald, 1983: 4).

²⁴ Selbstverständlich verringert dies nicht die Bedeutung der perzeptiven Distinktivität im Spanischen, die den Lautwandel der Sibilanten im Altspanischen erklärt.

²⁵ S. u.a. Trouvain (2004), Dankovičová (2001), Zellner (1998); auch Weiss (2008) sowie Pfitzinger (2001) zu Untersuchungen mit Literaturüberblick.

Pfitzingers „perzipierte lokale Sprechrate“ (s. Kap. 5) wird in Fenstern von 625 ms Länge gemessen.²⁶ Als globales Sprechtempo werden Durchschnittswerte bezeichnet, die für längere Zeiträume, z.B. ganze Äußerungen, erhoben werden und repräsentativ für Sprecher oder Sprachen verwendet werden.

2.3.1 Faktoren, die das lokale Sprechtempo beeinflussen können

Position:

Ein Auslöser für derartige Veränderungen kann die an Grenzen von Konstituenten produzierte finale Dehnung sein. Diese kann unterschiedliche Ausmaße annehmen und nur den letzten Silbenreim bis hin zum ganzen Wort erfassen²⁷. Grenzschnale werden mit hörerorientierten Prozessen, die zum Dekodieren von Nachrichten dienen, in Zusammenhang gebracht (z.B. Dankovičová, 2003). Auch die Informationsstruktur im Diskurs wird oft hiermit verbunden: Am Phrasenanfang stehen alte Informationen, die mit einer hohen AG geliefert werden, wohingegen am Phrasenende neue Information mit einer niedrigeren AG geliefert wird (Thema-Rhema-Struktur; z.B. Hanson, 2003). Gleichzeitig kann nicht ausgeschlossen werden, dass Verlangsamung und/oder Dehnung mit der Planung zusammenhängen (eine niedrigere AG verschafft mehr Zeit, um folgende Inhalte vorzubereiten, vgl. Fon, 1999). Sowohl das Ausmaß als auch die betroffenen Konstituenten unterscheiden sich zwischen Sprachen (Delattre, 1966; Lehiste, 1972).²⁸

Neben finaler wird auch auf initiale Dehnung hingewiesen, u.a. von Fant et al. (1991) und von Barbosa/Madureira (1999) für BP in Abhängigkeit vom Akzent der vorhergehenden gedehnten finalen Silbe.²⁹

Lautdauer:

Da Laute verschiedene inhärente Längen aufweisen, ist die durchschnittliche Lautdauer ständigen Veränderungen unterzogen. Die Artikulation von Vokalen wird als langsamer

²⁶ Die Betrachtung der Geschwindigkeit in einem zeitlichen Rahmen von 3 s Länge bezeichnet er als nicht mehr lokal (Pfitzinger, 2001: 124). Außerdem versteht Pfitzinger unter der lokalen Geschwindigkeit nicht einen festen Wert, sondern die Erhebung in regelmäßigen Abständen (etwa alle 20 ms) eines sich ständig verändernden Wertes, der über einen längeren Abschnitt gemessen dann eine lokale Sprechgeschwindigkeitskurve ergibt.

²⁷ Ein jüngerer Literaturüberblick zum Thema wird in Johnson/Martin (2001) gegeben.

²⁸ Über finale Dehnung hinaus kann in verschiedenen Sprachen eine allgemeine Verminderung des Tempos innerhalb von Konstituenten wie der Intonationsphrase festgestellt werden (Dankovičová, 1999, 2001, 2003; Hanson, 2003; Hakokari et al., 2007).

²⁹ Nach endbetonten Formen, d.h. nach Hauptakzent, weist der initiale Konsonant der Folgesilbe höhere Dauer auf; dieser Effekt ist über eine Satzgrenze hinweg kaum stärker als über eine Phrasengrenze.

und weniger kompliziert bezeichnet (Kent, 1983). Für den konsonantischen Bereich fasst Laver (1994: 434 f.) zusammen, dass labiale Konsonanten länger sind als Alveolare und Velare, allerdings scheinen diese Unterschiede in der inhärenten Länge nicht in allen Sprachen gleich ausgeprägt zu sein. Im vokalischen Bereich ging man davon aus, dass geschlossene Vokale wohl aufgrund des geringeren Kieferabstandes und der damit verbundenen kürzeren Artikulationsbewegung weniger Zeit in Anspruch nehmen als offene Vokale, für die der Kiefer weiter auslenkt (u.a. Marín Gálvez, 1994 für Spanisch; aber vgl. Lindblom et al., 1987). Solé/Ohala (2010) weisen nach, dass in manchen Sprachen (in der Studie Katalanisch und amerikanisches Englisch) diese sekundären Eigenschaften von Sprechern und Hörern zur Unterscheidung ähnlicher Vokale, die einen geringen Unterschied im Öffnungsgrad aufweisen (/e/und /ɛ/), herangezogen werden. In anderen Sprachen dagegen (hier dem Japanischen) stellt die inhärente Länge einen auf die artikulatorische Beschaffenheit zurückführbaren Nebeneffekt dar, der nicht direkt zur Unterscheidung verwendet wird, da nur hier der Unterschied bei Veränderung der SG nicht konstant bleibt (wohl aber bei phonologisch kontrastiven Lang- und Kurzvokalen). Escudero et al. (2009) stellen fest, dass im Portugiesischen die Längenunterschiede zwischen offenen und geschlossenen Vokalen stärker ausgeprägt sind als in anderen indoeuropäischen Sprachen. Hohe und häufige Schwankungen der lokalen AG sind in Sprachen zu erwarten, in denen phonologische Längenkontraste³⁰ vorliegen und in denen folglich Unterschiede in der Dauer der Segmente deutlich hörbar sein müssen, z.B. im Estnischen, das einen dreifachen Längenkontrast³¹ aufweist.³² In Sprachen wie dem Spanischen dagegen, das keinen Längenkontrast aufweist, können unter sonst gleichen Bedingungen tendenziell weniger derartige Veränderungen angenommen werden. Im Japanischen, das einen binären phonologischen Längenkontrast sowohl bei Konsonanten als auch bei Vokalen aufweist, scheint der tatsächliche, phonetisch realisierte Längenkontrast zwischen kurzen und langen Segmenten mit einem Verhältnis von 1:3 in der Dauer vergleichsweise stark ausgeprägt zu sein (Hirata/Whiton, 2005; Hirata, 2004; zitiert in Port, 2007).³³ Zwischen gespannten und ungespannten Vokalen zeigt Hakkarainen

³⁰ Diese können primär, aber auch sekundär sein, vgl. Kohler (1995: 169 f.) zur Vokallänge im Deutschen.

³¹ Vgl. aber Dufter (2003), der Einwände von Trubetzkoy (1962) und Lehiste (1997) zum dreifachen Längenkontrast des Estnischen anführt, jedoch weitere Beispiele nach Ladefoged/Maddieson (1996) für Sprachen mit derartigen Oppositionen gibt.

³² Duez (1993: 36) weist auf eine Untersuchung von Fox/Lehiste (1987) hin, nach der diese Längenunterschiede keinen Einfluss auf die Genauigkeit der Einschätzung der SG/AG haben: Estnische und US-Amerikanische Versuchsteilnehmer liefern vergleichbare Ergebnisse, wenn Sie Längenunterschiede von Plosiven beurteilen.

³³ Als Beispiel sei hier ja. はつびょう ([hap:^hjo:], de. Veröffentlichung) angeführt, wo unter gleichen Bedin-

(1995: 58) unterschiedliche Dauerunterschiede für Englisch, Deutsch, Schwedisch und andere Sprachen. Im Deutschen ist dabei der quantitative Unterschied am schwächsten ausgeprägt.

Lautliche Umgebung:

Stimmhafte Konsonanten lösen eine Längung des vorausgehenden Vokals aus, was im Englischen den Effekt der Überkreuzung phonetischer und phonologischer Längen zur Folge hat, wenn ungespanntes [ɪ] in „bid“ länger ist als gespanntes [i] in „beet“ (Calvert 1992: 229). Fischer-Jørgensen (1964; zitiert in Lehiste, 1970) erklärt außerdem, dass Vokale umso mehr Zeit in Anspruch nehmen, je weiter die Strecke ist, die die Zunge aus der Ausgangsposition zur Zielposition eines mit der Zunge gebildeten folgenden Konsonanten zurücklegen muss. Vor [d] werden bei dänischen Sprechern längere Werte für [u] gemessen als vor [g], vor [b] ist der Vokal noch kürzer, da der folgende Konsonant mit anderen Artikulatoren gebildet wird. Lenneberg (1986: 124) misst in en. „obtain“ eine deutlich längere Spanne für [t] als in „optimal“ (50 ms > 20 ms), so dass auch in Konsonantenhäufungen von gegenseitiger Beeinflussung der einzelnen Segmente bezüglich ihrer Dauer auszugehen ist.

Betonung:

Phonologisch akzentuierte Elemente, betonte Silben im Wort, aber auch auf Satzebene, werden in einigen Sprachen mit einer geringeren AG markiert. In Sprachen, in denen nur schwere Silben akzentfähig sind, kann eine längere Silbendauer schon durch eine höhere Anzahl von Segmenten (nämlich einen schließenden Konsonanten oder gebrochenen Vokal) oder auch einen langen Monophthong erreicht werden, was das Hervortreten betonbarer Silben begünstigt (s. Restle/Vennemann, 2001 zu verschiedenen Silbengewichtssystemen). Eine phonetische Längung betonter Silben kann aber auch im Spanischen beobachtet werden (s. Marín Gálvez, 1994), einer Sprache, in der das Silbengewicht keine Rolle spielt und für die lange Zeit tendenziell einheitlich lange Silben angenommen wurden.

Relevanz:

Quené (2008: 1105) verweist auf Zwaardemaker/Eijkman (1928) und Nooteboom/Eefting (1994) und hält fest: „important or unpredictable portions are typically spoken at a relatively slower rate“. Selbstverständlich können nicht nur einzelne Silben oder Wörter,

gungen eine längere Wortdauer zu erwarten ist als im annähernd vergleichbaren en. „happy“, das keine phonologisch langen Segmente aufweist.

sondern auch ganze Satzteile oder Sätze betont und durch eine niedrige AG markiert werden. Auer et al. (1999) bringen eine hohe Silbengeschwindigkeit bei gleichzeitig niedriger Betonungsdichte mit niedriger Relevanz in Verbindung. Umgekehrt markieren niedrige Silbengeschwindigkeit und hohe Dichte relevante Abschnitte im Gespräch (bzw. in der Äußerung). Das Sprechtempo wird damit als Beitrag zur Informationsorganisation („information management“) bezeichnet.

Semantischer Gehalt:

Phonologische Wörter mit lexikalischer Information, solche mit lexikalischem Inhalt und grammatischer Funktion und solche mit ausschließlich grammatischem Inhalt verhalten sich unterschiedlich, was die lokale AG anbelangt (Dankovičová, 1999). Dabei werden grammatische Wörter am schnellsten und lexikalische am langsamsten artikuliert. Die Ergebnisse beziehen sich auf Tschechisch und Englisch und entsprechen Feststellungen in Hegedüs (1957) für Ungarisch. Maclay/Osgood (1959) finden außerdem heraus, dass vor lexikalischen Wörtern Pausen wahrscheinlicher sind als vor grammatischen, was wiederum die lokale SG beeinflusst. In diesem Kontext soll noch ein weiterer Aspekt, der die lokale SG v.a. durch die Position von Pausen beeinflussen kann, erwähnt werden. Stanienda (2002: 42) weist auf die Gebrauchshäufigkeit von Wörtern hin: Pausen sind vor seltener gebrauchten Wörtern wahrscheinlicher als vor hochfrequenten. Pickett (1980: 95) erwähnt Ergebnisse von Umeda (1975), aus denen hervorgeht, dass die Gebrauchshäufigkeit eines Wortes innerhalb einer Unterhaltung die Dauer des betonten Vokals beeinflusst: „[W]hen an initially unpredictable word, *father*, occurred repeatedly during a discourse, the main vowel [ɑ] was much longer the first few times the word occurred than later in the discourse“. Dies kann als Kontextualisierung neuer Information interpretiert werden. Auch Lenneberg (1986: 117) betont die Frequenz von Wörtern als Einfluss auf die Geschwindigkeit, was er von einem artikulatorischen Standpunkt aus erklärt („Man muß bestimmte Wörter einige Male gebraucht haben, ehe man sie mühelos aussprechen kann“).

Diskursstrukturierung:

Innerhalb von Gesprächsbeiträgen werden häufig Zwischenbemerkungen durch eine veränderte SG kontextualisiert (Uhmann 1997: 205 ff.). Lokale Veränderungen dienen auch als *Turnhaltesignal* (Selting 1995: 98 ff.). Der Begriff „Durchhecheln“ wird verwendet, um die deutliche Erhöhung der SG zu bezeichnen, mit der ein Sprecher vermeiden möchte, dass ein anderer Sprecher vorzeitig das Wort ergreift. Bei Schegloff (1982; 1996) wird dieses Phänomen als „rush through“ bezeichnet und der Zusammenhang mit der Atmung erwähnt: Eine Atempause wird dabei vom Ende einer (syntaktischen,

intonatorischen und/oder pragmatischen) Einheit in die anschließende Einheit desselben Sprechers verschoben (s. auch Ford/Thompson, 1996). Couper-Kuhlen (1993) zeigt außerdem, wie die SG zur Kontextualisierung von *repair*-Sequenzen eingesetzt wird.

Länge der Konstituente:

Je mehr Silben im Wort enthalten sind, desto kürzer fallen diese einzelnen Silben aus (vgl. Menzerathsche Gesetze). Lindblom (1983: 226 f.) unterstreicht diesen bereits in Stetson (1951)³⁴ erwähnten Effekt mit Verweis auf Lyberg auch für bedeutungslose Logatome im Schwedischen: „This shortening, a characteristic of words in many languages, increases with the number of following syllables (Lyberg, 1981)“. Das gilt auch für die sogenannten silbenzählenden³⁵ Sprachen. Aus den Daten in Almeida (1997) geht für das in Tenerifa gesprochene Spanisch eine (allerdings nicht lineare) Kürzung der Silben in längeren Füßen hervor. Auch auf der Ebene des Satzes und der artikulatorischen Phrase können derartige Kompressionen festgestellt werden.³⁶ Zur kompensatorischen Kürzung im segmentalen Bereich in komplexeren Silben zeigen Untersuchungen (darunter Menzerath/de Oleza, 1928, zum Spanischen und Munhall et al., 1992, zum Englischen), dass die Dauer einzelner Laute mit der Anzahl der Laute in der Silbe abnimmt.

Globales Sprechtempo:

Erhöhtes Sprechtempo geht oft mit Reduktionen des phonetischen Sprachmaterials einher. Der Grund für Reduktionen liegt nicht ausschließlich im Sprechtempo: „[R]eductions are related to speech rate, but also reflect the degree of effort the subject is willing to make. Fast rate is a sufficient cause, but not a necessary one“ (Shockey 1987: 223). Später erklärt Shockey genauer:

„[H]uman beings probably do not, in fact, ever speak so fast that it is impossible for them to realize all the phonetic distinctions which would be present in a slow version of any given utterance. [...] But, weighing the effort that would be needed to realize all these distinctions against loss of intelligibility, the phonology allows many reductions to occur.“ (ebd.)

In ihrer Untersuchung sind Versuchsteilnehmer dazu fähig, alle Artikulationsbewegungen auch in beschleunigten Versionen auszuführen (vgl. Kuehn/Moll, 1976), was dafür spricht, dass die durchschnittliche Zeit, die für eine Artikulationsbewegung zur Verfügung steht,

³⁴ Stetson bezieht sich auf S. 102 auf „studies of Rousselot, et al.“ (Rousselot/Précis, 1913).

³⁵ Obwohl eine Einteilung der Sprachen der Welt in akzent-, silben- und morenzählende Sprachen abgelehnt wird, werden die Begriffe in dieser Arbeit v.a. bei Verweisen auf bisherige Untersuchungen verwendet.

³⁶ Da auch ganze Äußerungen von diesem Effekt betroffen sein können, wird er weiter unten bei den Einflüssen auf die globale SG aufgeführt.

linear gekürzt werden kann, ohne dass dabei Artikulationsbewegungen entfallen *müssen*. Nach einer Untersuchung von Cutugno/Savy (1999) zu italienischer Spontansprache hängt die Anzahl der Reduktionen nicht direkt mit einer steigenden lokalen AG zusammen. Daneben verhindert eine Verminderung der Geschwindigkeit Reduktionen nicht.

Da der Längenunterschied zwischen gespannten und ungespannten Vokalen im Englischen (und anderen Sprachen) neben spektralen Informationen auf der relativen Dauer der Segmente beruht, beeinflusst das globale Sprechtempo die Wahrnehmung der Vokallänge. In Versuchen können Johnson/Strange (1982) für Englisch nachweisen, dass gespannte (lange) Vokale, wenn sie aus schneller gesprochenen Aufnahmen ausgeschnitten und in langsamer gesprochene Aufnahmen eingefügt werden, als ungespannte (kurze) Vokale identifiziert werden. In Gottfried et al. (1990) wird gezeigt, dass Hörern zum Unterscheiden der englischen Begriffe „bit“ und „beat“ sowie „bet“ und „bat“ relative zeitliche Informationen von Vokalen (in Abhängigkeit der AG) ausreichen, um diese korrekt zu identifizieren (vgl. Solé/Ohala, 2010). Das Gehör passt sich also der lautlichen Umgebung an und interpretiert die Länge der Vokale in Abhängigkeit der Länge der umgebenden Segmente.

Auf der Satzebene scheinen derartige Effekte auch nachweisbar zu sein. Banzina/Dilley (2010) manipulieren Wörter in informeller Umgangssprache des Russischen so, dass durch eine zeitliche Komprimierung unbetonter Vokale und umgebender Konsonanten Wörter entstehen, die auch existierenden Wörtern ohne phonologisch zugrunde liegenden Vokal an dieser Stelle entsprechen könnten, d.h. /stərana/ wird zu einem Wort komprimiert, dass als phonetische Realisierung von /strana/³⁷ interpretiert werden kann. Dabei entspricht die komprimierte Version des ursprünglichen, längeren Wortes nicht vollständig dem kürzeren Wort ohne Vokal. Die verwendeten Aufnahmen der Sprecher wurden erst am Ende längerer Durchläufe aufgenommen, um aufgrund der Ermüdung der Teilnehmer gezielt spektral reduzierte Äußerungen zu erlangen. Die manipulierten und nicht manipulierten Versionen der Wörter werden so mit zeitlich komprimierten und nicht komprimierten Kontexten kombiniert, dass sich fünf Stimuli ergeben: eine unveränderte Originalversion eines vollständigen Satzes, nicht manipulierter Kontext mit komprimiertem Wort, nicht manipulierte Wort in negativ komprimiertem Kontext, negativ komprimiertes Wort in nicht manipuliertem Kontext und nicht manipulierte Wort in komprimiertem Kontext. Von Muttersprachlern des Russischen (N=20) wird in achtzehn derartigen Satzpaaren, in denen jeweils das kürzere wie auch das längere Testwort einen grammatischen und

³⁷ Die Autoren geben keine Betonungen an.

inhaltlich sinnvollen Satz ergibt, tendenziell das kürzere Wort wahrgenommen, wenn das Testwort einen komprimierten Anteil am Kontext ausmacht: in 62% der Fälle, wenn das Zielwort komprimiert ist, bzw. in 74% der Fälle, wenn der restliche Satz gedehnt ist. Somit steht fest, dass lexikalische Einheiten unter Berücksichtigung zeitlicher Aspekte wahrgenommen und verarbeitet werden.

Darauf, dass die Geschwindigkeit der lautlichen Umgebung auch einen Einfluss auf die wahrgenommene Länge von Elementen hat, weist auch Bose (1994: 98) im Zusammenhang mit Pausen hin. Objektiv längere „gefüllte Pausen“ können in Abhängigkeit des Kontextes als kürzer im Vergleich zu objektiv kürzeren Pausen empfunden werden. Eine Pause von 0,63 s Länge kann in einer mit durchschnittlich 3,57 Silben/s produzierten Äußerung als länger wahrgenommen werden im Vergleich mit einer 0,83 s langen Pause, die in einer 6,36 Silben/s schnellen Äußerung produziert wird (hierzu auch Meinhold, 1967).

Geschlecht des Sprechers:

In Stölten/Engstrand (2002) werden bei Frauen längere VOTs und längere Behauchung gemessen als bei männlichen Sprechern. Größere zeitliche Ausdehnungen im segmentalen Bereich erklären ein geringeres durchschnittliches Tempo bei den weiblichen Sprechern. Zu ähnlichen Ergebnissen kommt Rao (2010), der längere Wörter, betonte Silben und finale Dehnung bei weiblichen Sprechern des Spanischen misst. Escudero et al. (2009) messen in BP und EP geringfügig längere Vokale bei weiblichen Sprechern. Transsexuelle Sprecher des Niederländischen produzieren signifikant längere isolierte Wörter, wenn sie in ihrer „neuen“ Identität als Frauen sprechen, im Vergleich zu ihrer ehemaligen Sprechweise als Männer (Gunzburger, 1995; sechs Sprecherinnen). Ob diese Unterschiede auf Stereotype, tatsächliche Unterschiede in der zeitlichen Organisation von Lauten oder andere Faktoren zurückzuführen sind, kann nicht eindeutig geklärt werden.

2.3.2 Faktoren, die das globale Sprechtempo beeinflussen können

Sprache:

Roach (1998) geht davon aus, dass keine systematischen Unterschiede bezüglich des Sprechtempos zwischen Sprachen bestehen, räumt aber ein, dass in einigen Sprachgemeinschaften langsames Sprechen eher akzeptiert wird. Grosjean vermutet in einer Zusammenfassung von vergleichenden Untersuchungen zwischen Gebärdensprache (American Sign Language, ASL) und gesprochenen Sprachen auf Bellugi/Fischer (1972) verweisend, dass die Anzahl von Propositionen pro Zeiteinheit gleich ist: „[A]ll languages

(oral or visual) will be characterized by the same rate of information output if such aspects as age of speaker, linguistic task, situation etc. are controlled“ (Grosjean 1980b: 309).

Abercrombie (1967: 96) vermutet andererseits, dass das Sprechtempo zu den bezeichnenden Eigenschaften einer Sprache zählt und Hammarström (1957: 288) spricht eine Vermutung bezüglich des Sprechtempos in Sprachfamilien aus: „[I]t is not too daring to assume that Germanic languages are spoken more slowly than the Latin tongues“, vgl. auch Moses (1964: 137) zu dieser Einschätzung. Gil (1986) bringt den phonologischen und syntaktischen Typus von Einzelsprachen mit dem typischen Sprechtempo in Verbindung: Als jambisch bezeichnet er Sprachen, die komplexe Silbenstrukturen, silbenzählenden Rhythmus und ein niedrigeres Sprechtempo (hier: Silbengeschwindigkeit) aufweisen. Ihnen stehen trochäische Sprachen mit einfachen Silbenstrukturen, akzentzählendem Rhythmus und höherem Sprechtempo gegenüber. Dellwo (2010) vermutet, entgegen Dankovičová (2003), dass der phonologische Rhythmustyp einer Sprache von der CV-Geschwindigkeit³⁸ abgeleitet werden kann.

Geschlecht des Sprechers:

Jacewicz et al. (2009) berichten von mehreren Ergebnissen bezüglich geschlechtsspezifischer Unterschiede im Sprechtempo: Laut den Untersuchungen von Byrd (1994) und Yuan et al. (2006) sprechen unter US-Amerikanern Männer etwas schneller als Frauen (vgl. auch Kendall, 2009). Auch Jacewicz et al. (2009) stellen in ihrer Untersuchung fest, dass Männer geringfügig (3,3%) aber signifikant schneller sprechen und schränken das Ergebnis auf vorgelesene Texte ein. Yuan et al. (2006) messen die SGs von chinesischen Sprechern³⁹ mit dem gleichen Ergebnis. In Verhoeven et al. (2004) ist dieser Unterschied zwischen den Geschlechtern etwas größer (6%; s. auch Quené, 2008).

Kowal/O’Connell (1980) fassen Ergebnisse zusammen, aus denen unter Grundschulern der zweiten und vierten Klasse die Jungen als langsamere Leser hervorgehen. Sie produzieren beim Lesen mehr und beim Nacherzählen von Comics längere ungefüllte Pausen als ihre Mitschülerinnen. Möglicherweise kann dieser Unterschied jedoch auf Unterschiede in der Entwicklung zurückzuführen sein.

³⁸ Dabei werden konsonantische und vokalische Intervalle gezählt, z.B. weist es. „casa“ die gleiche Anzahl an CV-Intervallen auf wie de. „Sprache“, nämlich jeweils zwei.

³⁹ Die Autoren vergleichen diese mit den Werten für amerikanisches Englisch. Dabei zählen sie im Englischen Wörter/min und Wörter/art. Phrase, im Chinesischen dagegen „Zeichen“ (also streng genommen Morpheme) anstelle von Wörtern.

In einer Untersuchung von Stanienda (2002) zum Pausenverhalten und „flüssigem Sprechen“ (L1 und L2) weisen Frauen in ihrer L1 eine höhere SG ($3,91 > 3,35$ Silben/s), dafür eine niedrigere AG ($5,07 < 5,22$ Silben/s) als Männer auf, d.h. sie machen deutlich weniger Pausen. Sie fasst abschließend zusammen, dass Frauen die flüssigeren L2-Sprecher sind, was die Untersuchungen jüngerer Sprecher bestätigt.

Alter:

Kowal et al. (1983) bezeichnen das Alter als eine der offensichtlichsten Variablen temporeller Phänomene gesprochener Sprache und warnen vor Pauschalisierungen über eine Sprechsituation hinaus. Jacewicz et al. (2009)⁴⁰ messen niedrigere SGs bei älteren Sprechern und verweisen auf physiologische Erklärungen in Ramig (1983). Aus Stölten/Engstrand (2002) geht hervor, dass im lokalen Bereich ältere Sprecher nur teilweise längere Segmente produzieren. Es wird eine längere Behauchung bei einem von zwei Testwörtern gemessen; die Zeitspanne der VOT ist dagegen bei jüngeren Sprechern länger. Die niedrigere SG basiert offensichtlich nicht auf temporellen Unterschieden im segmentalen Bereich sondern manifestiert sich womöglich in der Pausenstruktur deutlicher.

Dialekt:

Hinweise auf dialektale Unterschiede im Sprechtempo sind zahlreich. Hewlett/Rendall (1998) untersuchen Sprecher von den Orkney-Inseln und aus Edinburgh. Zwischen den Gruppen können beim Vorlesen keine signifikanten Unterschiede in der SG festgestellt werden. Signifikant ist dagegen ein Unterschied von 0,5 Silben/s in der AG bei Spontansprache. Die Sprecher des ländlichen Dialektes produzieren bei spontanem Sprechen fünf Prozent mehr Pausen.⁴¹

Jacewicz et al. (2009) zitieren Ergebnisse von Ray/Zahn (1990), die keine signifikanten Unterschiede zwischen Sprechern aus dem Nordwesten, Südwesten und mittleren Westen der USA finden, und von Byrd (1994), die Südstaatler als langsamere Sprecher im Vergleich mit Nordstaatlern ausmacht. In ihrer Untersuchung stellen die Autoren fest, dass Sprecher aus Wisconsin 8% (Vorlesen) bis 12,5% (Spontansprache) schneller sprechen als Sprecher aus North Carolina (statistisch signifikant).

Levine et al. (1980) messen die SG in Silben/s von Postangestellten, die in Columbus, Ohio, durchschnittlich schneller sprechen als in Los Angeles (s. aber Kendall, 2009, unten).

⁴⁰ Sowie die darin zitierten Arbeiten von Quené (2008), Verhoeven et al. (2004), Yuan et al. (2006), und Smith (1987).

⁴¹ S. entsprechende Vermutungen in Oksaar (1988: 53) zu niedrigem Tempo bei ländlichen Sprechern.

Die Studie scheint gängige Stereotype in den USA zu bestätigen, leider werden keine Werte (nur Rangordnungen) angegeben.

In Kendall (2009: 143) wird die Herkunftsregion des Sprechers als signifikante Variable des Sprechtempos dargestellt. Seine Zusammenfassung lautet: „Speakers from Texas, Southern N[orth] C[arolina], Ohio, and Eastern NC favor faster speech rates, whereas speakers from Central NC, Western NC, and Washington, DC favor slower speech rates“. Die Auswertungen belegen Unterschiede innerhalb der Region North Carolina, die in Jacewicz et al. (2009) nicht erkennbar sind. Kendall erklärt:

„I do not think that there is something necessarily special about Ohio or Ohioan English that causes those speakers to have shorter pauses and higher speech rates [gemeint: AGs] than most of the speakers of the other regions. Instead, I think that the results, as far as the category of “region” is concerned, indicate that pause duration and speech rate, are – in addition to the outcomes and indicators of cognitive processes – the results of both sociohistorical processes and localized practices, as are so many aspects of dialect.“ (Kendall 2009: 166)

Verhoeven et al. (2004) weisen einen Unterschied von 16% zwischen in den Niederlanden und Belgien gesprochenem Niederländisch nach. Innerhalb der Varietäten der Sprache werden dialektale Unterschiede festgestellt, die meist deutlich geringer und teilweise nicht signifikant sind. Die Autoren vermuten, dass die Verwendung des Niederländischen in Belgien seitens der Sprecher mit Eigenschaften einer Distanzsprache in Verbindung gebracht wird und Unterschiede in der SG hierauf zurückzuführen sein könnten. Quené (2008) bestätigt die Ergebnisse.

Leemann/Siebenhaar (2007) finden, dass Sprecher des Berner Dialekts ein niedrigeres Sprechtempo als Sprecher eines walliser Dialekts aufweisen, das sich v.a. an längeren Vokalen (statistisch signifikant) und einer stärkeren finalen Längung festmachen lässt. Offensichtlich bestehen zwischen Dialekten signifikante, wenn auch geringe Unterschiede im Sprechtempo, die schon auf lokaler Ebene zu beobachten sind.⁴² Im Vergleich der drei Standardvarietäten des Deutschen (in Deutschland, der Schweiz und Österreich) stellt Ulbrich (2005) fest, dass bezüglich der AG keine signifikanten Unterschiede vorliegen, diese aber in Hinblick auf die SG feststellbar sind. Deutsche Sprecher produzieren weniger und kürzere Pausen und weisen damit eine höhere SG in verschiedenen Textsorten auf.

Robb et al. (2004) finden heraus, dass die neuseeländische Varietät des Englischen schneller gesprochen wird als die US-amerikanische. In Robb/Gillon (2007) wird

⁴² Da die wissenschaftlichen Beweise hierfür nicht umfangreich sind, wurde darauf verzichtet, den Einfluss des Dialektes auf die lokale SG in 2.3.1 als eigenen Punkt aufzuführen.

festgestellt, dass bei Kindern dieser Unterschied genau umgekehrt ausgeprägt ist. In der Studie artikulieren US-amerikanische Kinder außerdem mehr ungespannte Vokale als neuseeländische Kinder.

In der Untersuchung von Escudero et al. (2009) wird festgestellt, dass im BP längere Vokale produziert werden als im EP. Gleichzeitig wird ein Unterschied bezüglich des Tempos zwischen den Varietäten ausgeschlossen: Bei einer genaueren Untersuchung im segmentalen Bereich kann festgestellt werden, dass wortmediale Konsonanten im BP kürzer sind, also eine kompensatorische Kürzung von Konsonanten nach Vokalen vorliegt.

In Rebollo Couto (1997) zeichnet sich ein spanischer Sprecher durch ein deutlich höheres Sprechtempo aus, das etwa doppelt so hoch ist wie das des langsamsten Sprechers, der aus Panama stammt. Dies entspricht einem gängigen Stereotyp, Spanier sprächen schneller als Lateinamerikaner (und Lateinamerikaner aus dem karibischen Raum besonders langsam), das jedoch auf keiner wissenschaftlichen Basis fundiert.

Soziale Herkunft:

Genauere Forschungen zum Einfluss diastratischer Merkmale des Sprechers auf das Sprechtempo wurden bislang wenige durchgeführt. In Kowal/O'Connell (1980: 64) werden zwei relevante Untersuchungen erwähnt:

„Low-socioeconomic urban kindergarten children used much longer, but fewer silent pauses than middle- or high-socioeconomic urban children in telling a story. Remarkably, by second grade, school socialization processes eliminate all such differences; the lower-socioeconomic level has caught up with the others in these respects (Bassett, O'Connell & Monahan, 1977). A similar study in Spanish investigated rural lower class children and urban upper and lower class children in the second grade in Guatemala. The urban poor were strikingly different from the other two groups: Although they articulated faster, their silent pauses were twice as long (4.5 seconds on the average), though of the same frequency, as those of the other groups (Bassett & O'Connell, 1978). In general, low-socioeconomic urban children need more time to produce less speech than other groups in comparable situations.“

Behauptungen wie die von Horn/Lehnert (1954)⁴³, „[v]erschiedene Gesellschaftsschichten haben verschiedenes Sprechtempo“, können aufgrund mangelnder Beweise nicht bestätigt werden.

⁴³ Zitiert in Hegedüs (1957: 9).

Idiolekt:

Im Gegensatz zu Goldman Eisler (1968) hält Quené (2008: 1107) fest: „[W]ithin-speaker variance is indeed far larger than between-speaker variance in speech tempo“. Er weist auch darauf hin, dass Unterschiede zwischen einzelnen Sprechern in den Niederlanden stärker ausgeprägt sind als in Belgien. Deese (1984) meint, dass kaum Abweichungen von einem mittleren Sprechtempo im britischen Englisch bei einzelnen Sprechern vorliegen. Statistisch geht aus einigen Studien hervor, dass die Unterschiede zwischen einzelnen Sprechern signifikant sind, z.B. bei Koreman (2006b), der schnelle und langsame Sprecher des Deutschen untersucht, wobei die Zuordnung zu diesen Geschwindigkeitskategorien auf dem subjektiven Höreindruck beruht.

Register und Sprechsituation:

Das Sprechtempo variiert in gelesenen Texten tendenziell weniger als in frei gesprochenen (Bose 1994: 138). Goldman Eisler (1968) beschreibt soziale Interaktion und die Art der Redeaufgabe als Einflussfaktoren der Länge von verbalen Häsitationen und damit der SG. Der Faktor „Redeaufgabe“ wird von Kowal et al. (1983) gegenüber der Muttersprache als wichtigerer Faktor der SG betont; zwischen Redesituationen mit und ohne rhetorischer Intention unterscheidet Kowal (1991). Dabei kann die rhetorische Intention eine Verdopplung des Pausenanteils bewirken. Spontansprache stellt sich in einigen Untersuchungen als durch relativ niedrigere SG ausgezeichnet heraus (z.B. Heilmann/Künzel, 2001; Barik, 1977); Calvert (1992: 178) geht dagegen von einer höheren SG bei Spontansprache aus.

In Bariks Untersuchung (1977) schwanken die SGs in fünf unterschiedlichen Redeaufgaben im Englischen und Französischen zwischen 2,2 Silben/s (Nacherzählung einer Geschichte) und 3,5 Silben/s (Vortrag); die AG liegt zwischen 4,3 Silben/s (Vortrag) und 5,1 Silben/s (Nacherzählung).⁴⁴ Noch stärker fallen im Französischen Unterschiede im Pausenanteil aus (18,4% bei einem Vortrag bis 56,6% beim Nacherzählen einer Geschichte). Im Englischen sind beide Unterschiede etwas geringer ausgeprägt.

Ein Einfluss des Verhaltens des Interviewers auf die interviewte Person kann von Webb (1969) nachgewiesen werden (vgl. auch Kendall, 2009). In standardisierten Interviewsituationen, in denen Interviewer sich an verbales und nonverbales Sprachverhalten normierende Regeln nach Matarazzo (1962) halten, beeinflusst die SG der

⁴⁴ Angaben im Original in Silben/min.

Interviewer die der Interviewten. Gegenüber nicht standardisierten Situationen fällt außerdem bei den Versuchsteilnehmern die Varianz im Sprechtempo geringer aus.⁴⁵

Als weiterer Aspekt soll die Haltung gegenüber dem Gesprächspartner erwähnt werden. Mehrabian (1972: 12) erwähnt Untersuchungen von Rosenfeld (1966), nach der ein höheres Sprechtempo mit dem Versuch, positiv auf den Gesprächspartner zu wirken, einhergeht. Weitere derartige Untersuchungen werden in 6.2 behandelt.

Psychische / emotionale Verfassung:

Von Essen (1949: 318) merkt an: „Im Sprechtempo offenbart sich sowohl die allgemeine seelische Haltung wie auch die zufällig, d.h. augenblicksbedingte seelische Verfassung und Bewegung“. Weiterhin zitiert er auf S. 319 Noreen (1923), der ein langsames Tempo mit „Ruhe, Selbstbeherrschung, gründlicher Überlegung, Zaudern und Gleichgültigkeit“ in Verbindung bringt und bei hohem Tempo an „Enthusiasmus, warmes Interesse, Eifer, Gedankenlosigkeit usw.“ denkt. Die genannten Assoziationen dürften allgemein nachvollziehbar, können aber durchaus kulturabhängig sein. Wertende Fehlschlüsse sollten natürlich vermieden werden: Auch wenn ein langsamer Sprecher den Eindruck vermitteln mag, nach „gründlicher Überlegung“ zu sprechen, so muss die Äußerung eines schnelleren Sprechers nicht zwangsweise als unüberlegt empfunden werden.

Traunmüller (2004: 658) assoziiert in seiner Übersicht über Stimmkorrelate von Emotionen und Einstellungen hohes Sprechtempo mit Zorn und Frustration, Furcht und Nervosität sowie Anflehung und Schmeichelei. Mit niedrigem Tempo verbundene Emotionen sind Betrübnis und Resignation sowie Drohung und Verachtung. Eine veränderliche SG wird dagegen mit Zufriedenheit und Fröhlichkeit verbunden. Bei Bergmann (1988: 1218) werden hohes Sprechtempo mit Freude, Ärger, Furcht und Gleichgültigkeit, niedriges Sprechtempo mit Verachtung, Langeweile und Traurigkeit verbunden,⁴⁶ wobei er selbst darauf aufmerksam macht, wie wenig Übereinstimmung in derartigen Assoziationen besteht.

Länge der Äußerung:

Kendall (2009) zeigt. Dass die Länge der Äußerung als hochsignifikanter Faktor andere Faktoren in den Hintergrund drängt: So ergibt sich unter Berücksichtigung der Länge der

⁴⁵ Hierzu sind die Erklärungen der *Speech Accomodation Theory* (auch *Communication Accommodation Theory*) zu erwähnen, nach der unter bestimmten Umständen die Verwendung verbaler und nonverbaler Kommunikationsmittel an die des Gesprächspartners angepasst werden.

⁴⁶ Unter Verweis auf Scherer (1979b).

Aussage für die diatopische Herkunft der Sprecher, die sonst signifikant ist, kein signifikanter Effekt mehr. Geschlecht und ethnischer Hintergrund bleiben in Kendalls Auswertungen weiterhin signifikant. Kendall fasst zusammen:

„[S]peech rate [...] is principally impacted by utterance length. This process – commonly referred to as ‘anticipatory shortening’ – is likely a phonetic universal, whereby longer utterances are spoken at a higher rate due to physiological (namely, respiration) limitations.“ (Kendall 2009: 153 f.)

In Robb/Gillon (2007) werden Korrelationskoeffizienten zwischen der Länge der Äußerung und dem Tempo (SG sowie AG) von 0,92 bis 0,97 berechnet. Möglicherweise kann diese Variable widersprüchliche Ergebnisse anderer Untersuchungen erklären, wie Quené (2008: 1108) anmerkt:

„The reduction of the sex and age effects [nachdem der Faktor Phrasenlänge berücksichtigt wurde], and the significant phrase length effect, together suggest that the significant effects of sex and age observed in previous analyses may have been inflated by between-speaker effects on phrase length.“

Kultureller Hintergrund:

Von Stern (1900: 116) wird der „neapolitanische mit dem friesischen Fischer“ verglichen und Unterschiede, die „die zeitliche Struktur des ganzen Seelenlebens überhaupt“ betreffen, vermutet. Laut Lehtonen/Sajavaara (1985: 197) werden die Sprachen Deutsch von den Ostfriesen, Französisch in Belgien, Schweizerdeutsch in Bern und Zürich sowie Finnisch in Häme besonders langsam gesprochen. Die Formulierungen in einigen Untersuchungen zur SG lassen Unterschiede im Sprechtempo zwischen Sprachen (vielmehr: Völkern) eher als kulturelles Phänomen denn als linguistisches erscheinen, und es ist nicht immer deutlich, ob als Faktor des Tempos die Sprache, geografische Region, Subkultur oder Nation gemeint ist.⁴⁷ Können Unterschiede zwischen Regionen als „kulturell“ bezeichnet werden? Von der Definition des Kulturbegriffs abhängig ist, ob der Unterschied zwischen den von Hewlett/Rendall (1998) untersuchten Gruppen, der von den Autoren unter dem Begriff „lifestyle“ behandelt wird, als „kulturell“ bezeichnet werden kann.

Kendall (2009) und Slembek (1993) gehen bei ihren Untersuchungen von Varietäten plurizentrischer Sprachen betont auf den kulturellen Aspekt ein. In Kendall (2009) stellt sich heraus, dass Lumbee-Indianer langsamer Englisch sprechen als andere US-

⁴⁷ „Franzosen, Italiener oder Griechen [...] sprechen im allgemeinen bei weitem schneller als Deutsche, Ungarn oder Türken mit ähnlicher Bildung“ (Horger 1929: 29f.); „Sprache, Argumentation und Gebärde des südlichen Volkes sind schneller, gelöster, eindringlicher – die Verhaltnheit, „Haltung“, Beherrschtheit der spanischen Vornehmen betonter, fühlbarer, geradezu gewaltsamer“ (Flachskampf 1938: 218).

Amerikaner. Slembek (1993) untersucht schweizerische Sprecher des Deutschen und Französischen und stellt, auf Lehtonen (1986) verweisend, fest, dass sich die Art (Dauer und Häufigkeit) von Pausen je nach Kultur unterscheiden. Dabei stellt sich heraus, dass germanophone Nachrichtensprecher Deutschlands und der Schweiz durchschnittlich längere Pausen produzieren (jeweils 0,53 s) als frankophone Nachrichtensprecher Frankreichs (0,3 s) und der Schweiz (0,38 s). In den französischen Nachrichtensendungen werden doppelt so viele Pausen produziert (sechs pro 100 Silben) wie in den frankophonen schweizerischen Sendungen (drei pro 100 Silben), in den deutschen Sendungen etwas weniger (sechs pro 100 Silben) als in den germanophonen schweizerischen Sendungen (acht pro 100 Silben). Der Pausenanteil liegt bei den germanophonen Schweizern am höchsten (4,27 s pro 100 Silben); hier wird auch die niedrigste SG gemessen (4,2 Silben/s). Damit können Unterschiede zwischen Sprachen (Pausendauer) als auch zwischen Kulturen (Pausenhäufigkeit und SG) festgehalten werden, aber auch Gemeinsamkeiten (Pausenhäufigkeit in deutschen und französischen Sendungen). Problematisch ist dabei die fehlende Definition „kultureller Unterschiede“. Ein sehr deutlicher Unterschied besteht in der Pausenhäufigkeit zwischen frankophonen und germanophonen Nachrichtensprechern in der Schweiz. Es wird nicht näher dargestellt, ob der kulturelle Hintergrund germanophoner Schweizer größere Ähnlichkeiten zu dem Deutschen oder frankophoner Schweizer aufweist und nicht erklärt, ob unterschiedliche Kulturen für Sprachgemeinschaften, Nationen, oder andere gesellschaftliche Formen angenommen werden.

Ein Hinweis darauf, dass nicht die Sprache, sondern die Sprechgewohnheiten in Sprachgemeinschaften das Tempo bestimmen, wird von Yuan et al. (2006) geliefert, die feststellen, dass Japaner deutlich langsamer Englisch sprechen als alle anderen untersuchten Nationalitäten. Außerdem sprechen Muttersprachler des Französischen und des Spanischen Englisch als L2 mit einer höheren SG (Wörter/min) als Muttersprachler des Englischen. Sprachgewohnheiten in der L1 werden also in die L2 „transferiert“, wovon auch das Sprechtempo betroffen ist.

Kulturelle Unterschiede bezüglich der Toleranz, Wahrnehmung und Interpretation von Stille und Pausen im Diskurs werden in den Untersuchungen von Morsbach (1988) zum Japanischen, von Basso (1972) zu den Apache-Indianern und Haggan (1973) zum Arabischen in Kairo dargestellt.

Weitere Aspekte:

Trouvain (2004) führt neben den genannten Faktoren u.a. die folgenden Punkte an, die hier bisher unberücksichtigt blieben: Sprechfertigkeit, Stress, Beeinträchtigungen des Hör-

und/oder Sprechvermögens, Gesprächsplanung und Informationsstruktur. Die Aspekte „Gesprächsplanung“ und „Informationsstruktur“ werden hier als diskursrelevante Aspekte betrachtet und daher nicht einzeln aufgeführt. Beeinträchtigungen der Sprache und Fertigkeiten in Zweit- und Fremdsprachen werden in dieser Arbeit nicht weiter untersucht. Der Aspekt „Stress“ wird als ein die emotionale Verfassung des Sprechers beeinflussender Faktor verstanden und daher ebenfalls nicht gesondert besprochen. Sicherlich können noch weitere Faktoren des Sprechtempos angenommen werden, die in bisherigen Studien nicht berücksichtigt wurden.

Lenneberg (1986) verweist auf Guiot et al. (1961), die eine Veränderung der SG bei elektrischer Reizung eines bestimmten Bereiches im Thalamus feststellen. Wenn man davon ausgeht, dass ein bestimmter Bereich im Gehirn die Steuerung der SG übernimmt, sind keine Unterschiede zwischen Sprachen zu erwarten, die als physisch bedingt belegt werden können. Kulturelle Normen können jedoch durchaus dazu führen, dass in Gesellschaften Fertigkeiten (z.B. schnelle Artikulation) unterschiedlich ausgebildet und eingesetzt werden. Die Verwendung der Fertigkeiten kann wiederum von den vielen weiteren genannten Aspekten beeinflusst werden.

2.4 Pausen

In Untersuchungen zu Pausen wird häufig der 1965 von Tosi eingeführte und 1980 von O'Connell/Kowal diskutierte Begriff „Pausology“ verwendet. In diesem Abschnitt soll nicht die umfangreiche Pausenforschung dargestellt werden, die seitdem viele Änderungen und Erweiterungen erfahren hat; es sollen lediglich verschiedene Beschaffenheiten und Funktionen von Pausen vorgestellt werden. Daneben wird auf die Wahrnehmung von Pausen und deren Zusammenhang mit dem Sprechtempo eingegangen. Zu Beginn werden einige grundlegende Unterscheidungen beschrieben, die für die empirische Studie im vierten Kapitel von Bedeutung sein werden.⁴⁸

2.4.1 Arten von Pausen

Zunächst können sogenannte „gefüllte“ von ungefüllten Pausen unterschieden werden. „Gefüllte“ Pausen beinhalten irgendeine Art von artikulatorischer Aktivität. Füllwörter („nun“), Fülllaute („hm“) und Längungen („dannn...“) o.Ä. fallen in diese Kategorie, wohingegen in ungefüllten (stillen) Pausen keine Artikulation stattfindet. Unterbrechungen

⁴⁸ Eine umfangreichere Pausenklassifizierung ist z.B. bei Ballmer (1980) zu finden.

im Sprachsignal können jedoch Bestandteile der Artikulation sein, z.B. die Phase des totalen Verschlusses bei Plosiven. Es wurde daher diskutiert, ab welcher Länge bei einer automatischen instrumentellen Erfassung von Pausen Unterbrechungen im phonetischen Signal als Pause gewertet werden sollen (der sog. „cut-off-point“). Kowal et al. (1983) liefern einen Überblick über bis dahin durchgeführte Untersuchungen zum Thema und die darin verwendeten Schwellenwerte, nach dem in der Mehrheit der Untersuchungen Werte zwischen 0,2 und 0,31 s verwendet werden.⁴⁹ Zu der Bezeichnung „gefüllte Pausen“ ist hinzuzufügen, dass diese nur als Pausen bezeichnet werden dürfen, wenn sie als Unterbrechung der Vermittlung der inhaltlichen Information einer Aussage verstanden werden: Der Informationsfluss wird dann als während ihrer Produktion unterbrochen oder verzögert verstanden. Dabei ist zum einen problematisch, dass diese Ansicht negativ konnotiert ist; zum anderen, dass diese vokalen Elemente durchaus Informationen (u.a. zur Einstellung und momentanen Verfassung des Sprechers, zum Diskursverlauf uvm., s.u.) vermitteln, die v.a. in spontanen Sprechsituationen (aber nicht nur hier) auf keinen Fall außer Acht gelassen werden dürfen. Der an sich widersprüchliche Begriff „gefüllte Pause“ wird in dieser Arbeit weitgehend vermieden und stattdessen „verbale Häsimationsmittel“ verwendet. Teilweise ist auch von „Verzögerungen“ die Rede, wobei hier keine Wertung impliziert ist. Außerdem kann nicht mit letztendlicher Sicherheit gesagt werden, ob tatsächlich eine zögerliche Haltung seitens des Sprechers vorhanden ist, diese vom Hörer fälschlicherweise interpretiert oder gar vom Sprecher vorgetäuscht wird.

Nach ihrer biologischen Funktion können stille Pausen in Atempausen und Nicht-Atempausen unterteilt werden. Atempausen dienen der biologischen Notwendigkeit der Atmung und sind im Normalfall frei von semantischem oder diskursrelevantem Gehalt.⁵⁰ Sie sind für gewöhnlich länger (Kowal, 1991; Butcher, 1981) und werden tendenziell an Grenzen syntaktischer oder phonologischer Einheiten produziert (Fónagy/Magdics, 1960; Butcher, 1981).

Weiterhin können kommunikative Pausen, die rhetorische Bedeutung oder diskursrelevante Information beinhalten sowie gliedernde Funktion aufweisen können, von kognitiven Pausen (z.B. zur Wortwahl) unterschieden werden. Zur kommunikativen Funktion von Pausen wurde festgestellt, dass verbale Häsimationsmittel häufiger verwendet werden, wenn die Gesprächspartner sich nicht sehen (Cook/Lallgee, 1967; Kasl/Mahl,

⁴⁹ S. auch Bose (1994: 71), die Schwellenwerte zwischen 40 und 300 ms erwähnt, und Hieke et al. (1983).

⁵⁰ Denkbar ist auch, dass beim Einatmen ingressive Laute produziert werden, etwa „ts“. In diesem Fall handelt es sich bei der Atempause gleichzeitig um eine Art verbale Häsimation.

1965). Es wird von einer Kompensation nonverbaler Mittel durch verbale Mittel ausgegangen, wenn der visuelle Kanal nicht zur Verfügung steht. Im Diskurs können verbale Häsimationsmittel auch als *Turnhaltesignale* genutzt werden (MacLay/Osgood, 1959). Nachdrücklicher Anspruch auf das Rederecht wird vermutlich durch weitere prosodische Besonderheiten wie das bereits erwähnte Durchhecheln und nonverbale Mittel ausgedrückt oder von ihnen begleitet. Die kommunikative Funktion von Pausen wird gegenüber der kognitiven Funktion von Hänni (1973) betont. Er weist nach, dass sich eine Störung der Pausen durch Fremdgeräusche und Präsentation von Aufnahmen nicht auf die Sprachproduktion auswirkt bzw. auswirken muss. Es scheint in diesem Fall eine „Aufmerksamkeitsschwankung“ vorhanden zu sein, da der Inhalt in Pausen präsentierter Stimuli von den Sprechern aufgenommen wird (ebd., S. 98). Hänni berichtet außerdem davon, dass manche Sprecher unbewusst Geräusche produzieren (er beschreibt diese als „gutturales, knarrendes Geräusch“, S. 103), über die Versuchsteilnehmer selbst erstaunt sind, wenn sie das von ihnen produzierte Sprachmaterial in Aufnahmen anhören. Bestimmte verbale Häsimationsmittel können damit als teilweise automatisiert bezeichnet werden. Der Pause kommt (zumindest bei Prosaesungen) auch eine rhythmische Funktion zu (s. Snell 1918), was als euphonische Erscheinung gelten kann.

Eine kognitive Funktion kommt Pausen zu, wenn sie z.B. die Rede bei der Wortwahl verzögern. Zu dieser Funktion stellt Goldman Eisler (1968) – im Gegensatz zu Paivio (1965) – fest, dass Häsimationsmittel bei abstrakterem Sprachmaterial nicht häufiger sind, also unabhängig von kognitiven Prozessen produziert werden können. Andererseits zeigen MacLay/Osgood (1959), dass verbale Häsimationsmittel häufiger vor Inhaltswörtern verwendet werden und stellen einen Zusammenhang in Aussicht. Da Kendon (1980[1977]) feststellt, dass die Gestik während Pausen nicht unterbrochen wird und er davon ausgeht, dass Gestik und verbale Sprachproduktion zwei Aspekte ein- und desselben kommunikativen Prozesses sind, sieht er einen Beweis dafür, dass während Zögerungspausen Enkodierung stattfindet (s. Wallbott, 2003a: 270) und diese somit kognitive Bedeutung haben.

Alle erwähnten Pausen können an unterschiedlichen Positionen der Äußerung auftreten. Goldman Eisler (1968) stellt fest, dass sich Spontansprache im Gegensatz zu vorbereitetem Lesen dadurch auszeichnet, dass sich eine Vielzahl der Pausen nicht der linguistischen Struktur anpasst, d.h. nicht an Grenzen syntaktischer Konstituenten, sondern innerhalb solcher Einheiten produziert wird. Im Allgemeinen wird zwischen internen (innerhalb

syntaktischer Konstituenten) und externen Pausen unterschieden, die (Haupt- oder Neben-) Sätze voneinander abtrennen.⁵¹

Die Häufigkeit und die Position verbaler Häsitationsmittel werden von einer Reihe von Faktoren beeinflusst. Dazu sei hier kurz die Sprechsituation erwähnt: Maclay/Osgood (1959) zeigen, dass mehr Zögerungen beobachtet werden, wenn Sprecher in Vorträgen durch Unterbrechungen unter Druck geraten. Lallge/Cook (1969) können diesen Effekt für dialogische Sprechsituationen (Diskussionen) nicht nachweisen. Neben der Vertrautheit mit den Gesprächsteilnehmern sind auch das Redeverhalten des einzelnen Sprechers und die Verfassung zu nennen (vgl. Boomer, 1963 und Mahl, 1965 zur Auffassung gefüllter Pausen als Anzeichen für Verunsicherung). Brubaker (1972) interpretiert die Häufigkeit von Häsitationsphänomenen in vorgelesenen Texten als Unsicherheit über die Erfüllung der Aufgabe, einen unbekannten Text vorzulesen. Als determinierende Variablen des Pausenverhaltens können grundsätzlich alle unter 2.3 genannten Faktoren angenommen werden. Duez (1982) unterstreicht z.B. den Faktor „Sprecher“ in ihrer Untersuchung französischer Politiker. Kendall (2009) stellt einen Unterschied zwischen Frauen und Männern fest, was die mittlere Pausenlänge stiller Pausen anbelangt. Auch der Faktor „diatopische Herkunft“ scheint in seiner Untersuchung von Bedeutung zu sein.

Eine bedeutungsunterscheidende Funktion rechnen der Pause u.a. Quilis (1999) und Drommel (1974) zu. Quilis führt Beispiele an, die verdeutlichen, dass z.B. notwendige und nicht-notwendige Relativsätze anhand von Pausen unterschieden werden können.⁵² Drommel (1974: 39) unterscheidet Pausen mit den Merkmalen [\pm flüssig] und [\pm syntaktisch]. Dabei geht er von verbindenden T-Pausen (Transinformationspausen, die vom Sprecher intendiert sind und zur „kommunikativen Instruktion“ dienen) und abgrenzenden D-Pausen (Dissipationspausen, mit kognitiver Bedeutung) aus, die jeweils phonologisch die Merkmale [$+$ flüssig] bzw. [$-$ flüssig] tragen.

⁵¹ Goldman Eisler (1968: 13) verwendet die Begriffe „grammatical“ und „non-grammatical breaks“, Meinhold (1967: 107) „Satzpausen“ (innerhalb von Hypotaxen) und „Satzzwischenpausen“ (zwischen diesen), Stanienda (2002: 43) verwendet für „grammatikalisch motivierte Pausen“, die syntaktische Information beinhalten, den Begriff „juncture pauses“ (unter Berufung auf Deese, 1980; Perales/Cenoz, 1996).

⁵² Quilis/Fernández (1992: 164): „los alumnos, que viven lejos, llegan tarde“ im Gegensatz zu „los alumnos que viven lejos llegan tarde“ (de. Sinngemäß „nur diejenigen, die weit entfernt wohnen“ vs. „alle, die übrigens weit entfernt wohnen“). In Quilis (1993: 417) werden größtenteils die gleichen Bedeutungen von Pausen und Beispiele dazu angeführt, jedoch in anderer Reihenfolge. Ein interessantes Beispiel ist hier: „No, necesitamos estudiar más“ gegenüber „No necesitamos estudiar más“ (de. „Nein, wir müssen nicht...“ vs. „Wir müssen nicht...“), was häufig von Lernern im Anfängerstadium beim Vorlesen v.a. bei der Verneinung auf Satzebene und der zusätzlichen Verneinung direkt vor dem Verb falsch analysiert und missverständlich produziert wird („No, no [Pause] necesitamos estudiar más“, de. „Nein, nein, wir müssen...“).

Zu den sog. „gefüllten Pausen“ soll hier noch ergänzt werden, dass auch Versprecher, Stottern und ähnliche Phänomene („disfluencies“) als solche bezeichnet werden. Wird die Schnelligkeit der Artikulation betrachtet, so ist ihre Berücksichtigung bei Erhebung der AZ sinnvoll, da sie zum artikulierten Sprachmaterial gehören. Sie können andererseits ausgeschlossen und somit wie Pausen behandelt werden, wenn argumentiert wird, dass die Schnelligkeit der Artikulationsorgane während der Ausführung motorischer Befehle zur Versprachlichung propositionalen Gehaltes untersucht wird (weitgehend unabhängig vom Enkodierungsprozess). Auf den Artikulationsablauf können diese Unterbrechungen unterschiedliche Auswirkungen haben, wie Shriberg (1999: 620) anhand von Koartikulationsmustern bei Versprechern im amerikanischen Englisch zeigt. Dabei können vier Fälle unterschieden werden: Die Artikulation endet in 88% der untersuchten Fälle vor der Unterbrechung bzw. vor dem Abbruch, und zwar ohne Transition, oder mit Transition zum auf die Wiederholung folgenden Wort (in neun Prozent der Fälle), teilweise mit Transition auf ein anderes, vermutlich ursprünglich intendiertes Wort (2%) und auch mit Transition zum Wort, das wiederholt wird (0,3%). Für die Sprachproduktion bedeutet dies, dass Abbrüche in verschiedenen Stadien mit unterschiedlich weit vorangeschrittener Planung stattfinden können.

2.4.2 Zur Wahrnehmung von Pausen

In dem von Meinhold (1967) untersuchten Sprachmaterial (von Berufssprechern vorgelesene Belletristik, Lyrik und Nachrichten) stellen sich Pausen zwischen geschlossenen syntaktischen Einheiten als signifikant länger heraus.⁵³ Des Weiteren ist die Streuung deutlich größer als bei satzinternen Pausen. Pausen werden aber nicht immer vom Hörer wahrgenommen. In der Wahrnehmung spiegeln sich die von Meinhold gefundenen Unterschiede wider, wie Butcher (1981) festhält: Pausen zwischen phonologischen Einheiten („tone groups“) müssen länger sein (220 ms), um von Hörern wahrgenommen zu werden; Pausen innerhalb dieser Einheiten werden schon ab einer Dauer von 80 ms wahrgenommen. Somit kann davon ausgegangen werden, dass Pausen an bestimmten Positionen erwartet und perzeptiv nicht (oder unbewusst) verarbeitet werden. Folglich kann eine Unterscheidung in zu erwartende und nicht zu erwartende Pausen getroffen werden. U.a. betont Nooteboom (1985) außerdem, dass Pausen wahrgenommen werden,

⁵³ Vgl. auch Brubaker (1972), Butcher (1981), Duez (1982), Bose (1994), Yang (2007), und Oliveira (2002) zu geschlossenen Einheiten auf der Textebene.

wo keine produziert werden.⁵⁴ Duez (1993) verwendet für diesen Fall den Begriff „subjektive Pause“. Neben der Dauer von phonetischen Segmenten spielen nach Duez' Untersuchung auch die Intensität sowie der Tonhöhenverlauf eine wenn auch untergeordnete Rolle bei der Pausenwahrnehmung. Kowal/O'Connell (2003) fassen zusammen, wie häufig Pausen nicht wahrgenommen werden: Untrainierte Hörer nehmen 2,5% aller kurzen Pausen wahr,⁵⁵ bei trainierten Hörern sind es 52%.⁵⁶ Für die Untersuchung der AG bedeuten diese Ergebnisse, dass die perzeptive Erfassung von Pausen nicht ausreichend ist und instrumentelle messtechnische Verfahren unbedingt notwendig sind.

Nach Grammer (2004: 3468) können häufige Häsitationsphänomene wie Füllwörter und -laute den Eindruck erwecken, der Sprecher sei ängstlich oder gelangweilt; häufige stille Pausen können dagegen den Eindruck eines ängstlichen, verärgerten oder Verachtung empfindenden Sprechers erwecken (mit Verweis auf Lallge, 1971; Lallge/Cook, 1973). Diese Assoziationen können jedoch kulturgebunden sein.

2.4.3 Pausen im Sprachvergleich

Verbale Häsitationsmittel nehmen in verschiedenen Sprachen unterschiedliche Form an: Im Deutschen und Englischen können Fülllaute wie „ähm“ und „hm“ als häufig bezeichnet werden, im Spanischen scheinen Füllwörter (die ihre Grundbedeutung nicht zum Ausdruck bringen) wie „pues“ (lateinamerikanisches Spanisch häufiger: „este“) frequent zu sein und im Chinesischen werden oft ganze Phrasen wie 这个嘛 (zhè ge ma, wörtl. etwa „dieses also“) verwendet. Dabei tritt in den Füllwörtern im Spanischen (zumindest in den lateinamerikanischen Varietäten) und Chinesischen oft eine deutliche Dehnung (z.B. des letzten Vokals) hinzu.

Grosjean (1980a) fasst Unterschiede in der Pausenstruktur zwischen Englisch und Französisch in Interviews zusammen: Bei den „gefüllten“ Pausen weist das Englische deutlich mehr Fülllaute („uh“) auf als das Französische, was auf phonotaktische Verhältnisse zurückgeführt wird. Aufgrund der höheren Anzahl geschlossener Silben läge es im Englischen nahe, Häsitationsmittel in Form von Füllwörtern zu realisieren. Es wird keine überzeugende Erklärung angeführt, warum geschlossene Silben nicht gedehnt

⁵⁴ Dazu Nooteboom (1985: 247): „When the actual silent interval is omitted but there is prepausal lengthening, the auditory effect is very much the same“.

⁵⁵ Nach Carpenter/O'Connell (1988).

⁵⁶ Nach Kowal/O'Connell (1995).

werden können. Grundsätzlich scheint sowohl die Dehnung des Vokals als auch des auslautenden Konsonanten (oder, im Fall eines finalen Clusters, zumindest eines Bestandteils dieses Clusters, z.B. Sonoranten, teilweise aber auch Obstruenten) in geschlossenen Silben möglich, da diese auch von finaler Dehnung erfasst werden. Bezüglich des Pausenanteils wird festgestellt, dass dieser in den Sprachen vergleichbar ist (16,8% für Englisch und 15,5% für Französisch), im Englischen jedoch die Pausen geringfügig kürzer ausfallen (durchschnittlich 0,44 s im Vergleich zu 0,48 s) und häufiger sind. Als mögliche Gründe für längere artikulatorische Phrasen im Französischen (12 Silben gegenüber 9,5 im Englischen in Grosjean, 1980a und 6,22 Wörter im Französischen gegenüber 4,67 im Englischen in Grosjean/Deschamps, 1973) werden grammatische Strukturen angeführt, da viele Inhalte im Englischen in kürzerer Form ausgedrückt werden können.⁵⁷ Ein deutlicher Unterschied besteht in der Position von Pausen. 23% aller Pausen liegen im Englischen innerhalb der VP, wohingegen es im Französischen nur neun Prozent sind. Abschließend vermutet Grosjean, dass der Pausenanteil in unterschiedlichen gesprochenen Sprachen gleich hoch ausfallen müsste, syntaktische und morphologische Unterschiede jedoch unterschiedliche Positionen erlauben. Derartige Unterschiede sind durchaus plausibel: Im Japanischen und Türkischen, zwei agglutinierenden Sprachen, werden Konjunktionen als Suffixe an phonologische Wörter angehängt. Infolgedessen werden Pausen nach diesen Affixen produziert, wohingegen in indoeuropäischen Sprachen Pausen eher vor diesen Elementen, die meist freie syntaktische Einheiten darstellen, eingefügt werden. Vgl. tr. „ama“⁵⁸, ja. -と (-to), -も (-mo) vs. de. „aber“, „und“.

In Black et al. (1966) werden stille Pausen von Muttersprachlern der Sprachen Hindi, Spanisch und Japanisch untersucht, wobei ein signifikanter Unterschied weder in der Muttersprache noch wenn die Sprecher Englisch sprechen festgestellt werden kann (nach O'Connell, 1980). Vergleichbar lange durchschnittliche Pausendauer für amerikanisches Englisch und mexikanisches Spanisch werden von de Johnson et al. (1979) gemessen (694 und 733 ms). Beim untersuchten Material handelt es sich um Nacherzählungen von jeweils 45 Jugendlichen im Alter von 15 bis 18 Jahren von auditiv, visuell und audiovisuell präsentierten kurzen Geschichten. Die mexikanischen Jugendlichen produzieren längere

⁵⁷ Bsp.: „He came for his mail → il est venu prendre son courrier [...] *Do be careful* → *surtout faites bien attention* [...] the government's income policy → la politique des revenus du gouvernement [...] a shipyard → un chantier de construction naval [sic]“ (ebd., S. 46), wobei sich Gegenbeispiele finden lassen, z.B. „He came here in order to collect his mail“, „*do be careful*“ belegen die Verwendung zusätzlicher lexikalischer anstelle von prosodischen Mitteln zur Emphase.

⁵⁸ Zwar wird tr. „ama“ orthografisch vom vorangehenden Wort getrennt, es wird aber für gewöhnlich ohne vorangehende Pause angeschlossen. Da es betonbar ist, scheint es auch als eigenständiges phonologisches Wort verwendbar zu sein. I.d.R. wird die zweite Silbe mit einem Anstieg der Grundfrequenz verbunden.

Nacherzählungen (215 gegenüber 165 Silben, wobei die Gesamtzahl der artikulierten Wörter im Englischen höher liegt: 6297 gegenüber 5679; die durchschnittliche Anzahl von Silben pro Wort liegt bei 1,19 für Spanisch und 1,71 für Englisch). Im Englischen werden außerdem andere Häsitationsphänomene gefunden: Während hier Fülllaute frequent sind (z.B. „uh“, von 95% der US-Amerikaner aber nur 35% der Mexikaner verwendet), sind im Spanischen Füllwörter häufig (z.B. „este, pues, bueno“, bei 93% der Mexikaner aber nur 62% der US-Amerikaner aufgezeichnet). Die Verwendung bestimmter Häsitationsmittel scheint jedoch nicht ausschließlich sprachabhängig zu sein. Im europäischen Spanisch wird das Füllwort „este“ selten verwendet, hier sind „pues“ und „bueno“ häufiger; im Río-de-la-Plata-Gebiet fällt die Verwendung der Konjunktion „y“ auf. Den Varietäten des Spanischen gemeinsam ist, dass tendenziell grundsätzlich bedeutungsvolle Wörter als Häsitationsmittel verwendet werden. Ob sich Sprachen bezüglich der Pausenstruktur unterschiedlich verhalten, wird auch in Glukhov (1975) untersucht. Aus dem umfassenden Sprachmaterial geht hervor, dass die sechs untersuchten Sprachen sehr ähnliches Verhalten in Bezug auf die Häufigkeit von Pausen aufweisen. Am häufigsten werden Pausen von bis zu 150 ms produziert, längere Pausen werden mit zunehmender Dauer immer seltener. Dabei ergeben sich für die kürzesten Pausen, die eine Dauer von 50-150 ms aufweisen, Unterschiede zwischen den Sprachen. Im Portugiesischen werden etwa doppelt so viele dieser Pausen wie im Italienischen produziert. Dazwischen liegen, ohne größere Unterschiede, die Sprachen Deutsch, Französisch, Spanisch und Englisch (in aufsteigender Reihenfolge).

In der Untersuchung von Barik (1977) liegen fast identische gemittelte Werte vor (mittlere Pausenanteile für Französisch: 35,8%, für Englisch: 35,38%, eigene Berechnungen; durchschnittliche Pausenlänge für Französisch: 1,51 s, für Englisch: 1,4 s). In den fünf einzelnen Redeaufgaben sind jedoch deutliche Unterschiede im Bereich von bis zu 14% beim Pausenanteil (Spontansprache: Französisch > Englisch) und bis zu 0,61 s bei der durchschnittlichen Pausenlänge (formeller Vortrag: Englisch > Französisch) festzustellen. Der Anteil grammatischer Pausen liegt gemittelt bei 70,4% (Englisch) und 72,2% (Französisch), der größte Unterschied liegt mit 23,2% (Englisch > Französisch) bei der Redeaufgabe „informeller vorbereiteter Vortrag“ vor. Die Pausenstruktur (Anteil, Dauer und Position) variiert also der Situation entsprechend in den beiden Sprachen auf unterschiedliche Weise.

Sprachspezifische Unterschiede bei verbalen Häsitationsphänomenen sind in Bezug auf die verwendeten Häsitationsmittel nachweisbar. Der Pausenanteil scheint hauptsächlich in Abhängigkeit von situativen Faktoren zu variieren. Auch was die Position von Pausen

anbelangt lassen sich Unterschiede erkennen, die zumindest teilweise auf grammatische Strukturen der untersuchten Sprachen zurückzuführen sind.

2.4.4 Zusammenhänge von Sprechgeschwindigkeit, Artikulationsgeschwindigkeit und Pausen

Der Zusammenhang zwischen SG und Pausenverhalten ist offensichtlich: Mit steigendem Pausenanteil, der auf längere aber auch häufigere Pausen zurückgehen kann, sinkt die SG.⁵⁹ Umgekehrt wurden jedoch unterschiedliche Aussagen dazu gemacht, wie sich Veränderungen der SG in der Pausenstruktur bemerkbar machen.

Goldman Eisler (1968) betont die Bedeutung der Pausen für unterschiedliche Sprechtempi: „Variations in the overall speed of talking were found to be variations in the amount of pausing. [...] The rate of articulation based on vocal activity exclusively, on the other hand, was shown to be relatively invariant“ (Goldman Eisler 1968: 31). Auch Grosjean (1980a) führt die Veränderung der SG in erster Linie auf eine Reduktion der Anzahl der Pausen bei etwa konstanter AG und Dauer der Pausen zurück. Miller et al. (1984) dagegen finden in Interviews wesentliche Unterschiede in der AG innerhalb von Sätzen, die stärker als die Pausenstruktur Veränderungen der SG bewirken.

Zum Zusammenhang von AG und SG stellt Butcher (1981) fest, dass bei höherer SG Pausen zwischen Sätzen kürzer ausfallen, und gleichzeitig die AG zunimmt. In Boses spontansprachlichem Korpus besteht eine mittlere negative lineare Korrelation zwischen AG (Silben/s) und Pausenhäufigkeit (Anzahl der Pausen pro 100 Silben): $r=-0,63$ (Bose, 1994: 87). Damit wird ein Zusammenhang von AG und SG belegt, die gleichermaßen von Änderungen der Pausenstruktur betroffen zu sein scheinen. Die AG korreliert auch negativ mit dem Pausenanteil ($r=-0,78$).⁶⁰

Auf der anderen Seite stellt Kendall (2009) bei der Auswertung der Aufnahmen von 104 Sprechern innerhalb der Gruppe der Lumbee-Indianer (und nur hier) einen deutlichen linearen Zusammenhang zwischen AG und Pausendauer von $r^2=0,95$ fest – also einen negativen Zusammenhang von SG und AG. Er schlussfolgert:

⁵⁹ Vgl. aber Zellner (1998: 3159), in deren Untersuchung eines Sprechers des Französischen längere Pausen sowie häufigere Pausen (im Verhältnis 1:7) zusammen nur etwa 8% der zusätzlichen SZ von schnellem zu langsamem Tempo erklären.

⁶⁰ In dem Korpus korreliert der Pausenanteil sowohl mit der mittleren Pausendauer als auch mit der Pausenhäufigkeit stark (positiv).

„[S]ince speech rate, as analyzed here [gemeint: AG gemessen in Silben/s], is pause-exclusive, there is nothing mechanistic that connects these two measures. Instead, what we see here is the complex coordination of cognitive, physiological (i.e. articulatory and respiratory), and discourse processes.“ (Kendall 2009: 165)

Auch Barik (1977) misst hohe Werte für die AG bei niedriger SG (und umgekehrt); die nicht signifikanten Korrelationen betragen -0,7 (Englisch) und -0,8 (Französisch).

In Anbetracht der unterschiedlichen Ergebnisse muss von verschiedenen Sprechstrategien ausgegangen werden, in denen SG und AG positiv miteinander und beide negativ mit dem Pausenanteil (Häufigkeit und/oder Dauer) korrelieren, aber auch unabhängig voneinander variiert werden können. Im Gegensatz zur Pausenstruktur wird die Schnelligkeit der Artikulationsorgane nicht in erster Linie als von kognitiven Prozessen geprägt verstanden. Sie kann durchaus gezielt zu kompensatorischen Zwecken eingesetzt werden, um eine globale Konstanz der SG zu erreichen. Es gibt jedoch keinen Anlass, eine Abhängigkeit der AG von Veränderungen der SG und/oder des Pausenverhaltens anzunehmen.

2.5 Zusammenfassung

Hinter dem Begriff „Sprechtempo“ können sich viele unterschiedliche Vorstellungen verbergen. Zunächst ist die Unterscheidung von AG und SG grundlegend. Für die SG ist die Pausenstruktur von Bedeutung. Hier kann (mindestens) zwischen stillen Pausen und pausenähnlichen Abschnitten im Redefluss mit artikulatorischer Aktivität, für die die Bezeichnung „Häsitationsmittel“ vorgezogen wird, unterschieden werden. Außerdem wird zwischen internen und externen, zwischen Atem- und Nicht-Atempausen und kommunikativen und kognitiven Pausen unterschieden. Weiterhin muss berücksichtigt werden, dass für die Erhebung des Sprechtempos verschiedene Messgrößen unterschiedliche Sachverhalte darstellen.

Der Vergleich von Lauten, Silben, Wörtern, Morphemen oder anderen sprachlichen Elementen pro Zeiteinheit liefert unterschiedliche Bilder, die die Schnelligkeit der Artikulationsorgane oder der Übermittlung von Information beschreiben können. Es ist auch von Bedeutung, in welcher Form diese Elemente gezählt werden (tatsächlich realisiert oder kanonisch zugrunde liegend). Außerdem ist der zeitliche Rahmen von Messungen zu berücksichtigen und zwischen lokalem und globalem Tempo zu unterscheiden. Prinzipiell kann für das Sprechtempo ein Einfluss der Faktoren Sprache, Geschlecht, Alter, diatopische und diastratische Herkunft der Sprecher, Idiolekt, Sprechsituation, psychologisch-emotionale Verfassung des Sprechers, Länge der Aussage, kultureller

Hintergrund des Sprechers, Gegebenheiten im Diskurs und Besonderheiten auf lokaler Ebene (darunter finale Dehnung und inhärente Länge einzelner Segmente) nachgewiesen werden.

Im anschließenden Kapitel werden sprachvergleichende Untersuchungen des Sprechtempos vorgestellt. Außerdem werden, vom phonetisch-phonologischen Standpunkt aus betrachtet, zwei Phänomene thematisiert, die einen Zusammenhang mit dem Sprechtempo aufweisen. Es handelt sich um Koartikulation und dem, was allgemein als „Sprachrhythmus“ bezeichnet wird.

3 Phonetisch-phonologische Determinanten des Sprechtempos

3.1 Einleitung

Artikulatorisch betrachtet ist es vorstellbar, dass der phonologische Typ einer Sprache sowohl die Silben- als auch die Lautgeschwindigkeit beeinflusst. In Konsonantenhäufungen können die modifizierenden Artikulationsbewegungen einen relativ hohen artikulatorischen Aufwand erfordern. Ab einem bestimmten Verhältnis von Vokalen zu Konsonanten, das deutlich zu Gunsten der Konsonanten ausfällt, kann daher eine „bremsende Wirkung“ vermutet werden. Diese Bewegungen stellen Hindernisse dar, die den unbeeinträchtigten Bewegungsablauf vokalischer Gesten verlangsamen. Was den Zusammenhang von Tempo und Rhythmus anbelangt, so wird hier wie bei Fenk-Oczlon/Fenk (2006) davon ausgegangen, dass eine längere durchschnittliche Silbendauer, die mit einem hohen Konsonantenanteil einhergeht, neben einer niedrigen Silbengeschwindigkeit einen bestimmten wahrgenommenen Sprachrhythmus bewirkt (vgl. 4.4). Die bekannte Einteilung von Sprachen in Rhythmusklassen basiert auf dem Unterschied zu Sprachen, die einfachere Silben und folglich eine höhere Silbengeschwindigkeit aufweisen.

Das Kapitel beginnt mit einem Überblick über die bisherige Literatur zum Thema Sprechtempo im Sprachvergleich. Es wird sowohl auf Arbeiten eingegangen, die Gemeinsamkeiten aufdecken, als auch auf solche, in denen Unterschiede zwischen Sprachen aufgezeigt werden. Anschließend wird ein möglicher Zusammenhang zwischen Tempo und Koartikulation thematisiert. Ergebnisse sprachvergleichender Untersuchungen der Koartikulation werden in 3.3 vorgestellt. Im vierten Abschnitt des Kapitels werden Erkenntnisse der (phonologisch orientierten) Rhythmusdebatte vorgestellt. Die Rhythmusforschung ist deshalb relevant, weil sie unterschiedliche Silbenpräferenzen behandelt und diese einen Einfluss auf die durchschnittliche Silbendauer haben, die wiederum einen Aspekt des Sprechtempos darstellt. Damit wird zu den Ergebnissen der empirischen Studie übergeleitet.

3.2 Bisherige Forschungsergebnisse sprachvergleichender Untersuchungen zum Sprechtempo

Zwar stellt den Os (1988) keinen Unterschied in der AG in Silben/s zwischen Niederländisch und Italienisch fest (jeweils sechs männliche Sprecher lesen einen vergleichbaren Text vor, wobei die italienischen Aufnahmen signifikant länger sind). Unterschiede werden aber deutlich, wenn die AG in Lauten/s gemessen wird. Dann ist das Italienische signifikant langsamer als das Niederländische ($13,8 < 16,4$). Da im Niederländischen Konsonantencluster häufig, im Italienischen dagegen äußerst selten sind, entspricht dies der Prognose Gils (1986; s. 2.3.2) und steht im Gegensatz zur Hypothese dieser Arbeit, nach der bei einem hohen Konsonantenanteil eine niedrige Lautgeschwindigkeit zu erwarten ist.

Russo/Barry (2008) erhalten ähnliche Ergebnisse beim Vergleich deutscher und italienischer Sprecher. Deutsche produzieren zwischen 13,48 und 15,24 Laute/s bzw. 5,34 bis 6,08 Silben/s, und Italiener 8,21 bis 10,22 Laute/s bzw. 3,17 bis 4,77 Silben/s. Darüber, ob hierbei situative Gegebenheiten einen Einfluss haben, kann nur spekuliert werden. Die italienischen Aufnahmen stellen Wegbeschreibungen dar, die unter Zuhilfenahme von Karten spontan produziert werden; in den deutschen Aufnahmen handelt es sich um Gespräche, in denen Termine vereinbart und in Kalender eingetragen werden, wo möglicherweise ein Gewohnheitseffekt nicht ausgeschlossen werden kann. Die Sprechsituationen sind zumindest nicht gleich und gehen möglicherweise mit unterschiedlichen kognitiven Belastungen einher. Generell dürften im Italienischen Silben durchschnittlich etwas kürzer (da weniger komplex) sein, weshalb die niedrigere Silbengeschwindigkeit überraschend ist. In der Untersuchung zeichnen sich italienische Sprecher also tatsächlich durch ein niedrigeres Tempo aus.

Weitere Arbeiten, in denen bei Betrachtung eines Parameters des Sprechtempos kein Unterschied zwischen Sprachen festgestellt werden kann, sind Osser/Peng (1964) zu Japanisch und amerikanischem Englisch (untersucht: Laute/s), Carrió i Font/Ríos Mestre (1991) zu Spanisch und Katalanisch (kanonische Silben/s), Grosjean/Deschamp (1973; 1975) sowie Barik (1977) jeweils zu Französisch und Englisch (Silben/min).

De Johnson et al. (1979) dagegen belegen hochsignifikante Unterschiede zwischen amerikanischem Englisch und mexikanischem Spanisch. Die Silbengeschwindigkeit liegt bei spontanen Nacherzählungen im Spanischen höher (SG: $4,24 > 3,66$; AG: $6,08 > 5,00$). Des Weiteren sind die Nacherzählungen der mexikanischen Teilnehmer länger ($215 > 165$).

Silben). Leider werden die Informationsdichte und Genauigkeit der Nacherzählungen nicht untersucht, so dass nicht festgestellt werden kann, ob mexikanische Sprecher in den längeren Nacherzählungen mehr Informationen (detailliertere Nacherzählungen) liefern.

Weitere Unterschiede stellt Slembek (1993) bezüglich der SG fest. Sie misst 5,32 Silben/s bei französischen Sprechern; 5,04 bei französischsprachigen Schweizern; 4,84 bei deutschen Sprechern und 4,2 bei deutschsprachigen Schweizern. In Wörtern/s gemessen ergibt sich eine andere Rangfolge: 3,27 (frankophone Schweizer) > 3,15 (Franzosen) > 2,38 (Deutsche) > 1,77 (germanophone Schweizer).⁶¹ Der Umfang des Materials (sieben Aufnahmen, maximale Länge: 102 s) lässt vorerst keine Generalisierungen zu.

Kurze Ausschnitte von bis zu 22 s Dauer aus brasilianischen und spanischsprachigen Nachrichtensendungen (je drei weibliche und drei männliche Sprecher) werden in Rebollo Couto (1997) verglichen. Die spanischsprachigen Sendungen stammen von unterschiedlichen lateinamerikanischen und spanischen Sendern. Bei der Auswertung werden im BP stärkere Schwankungen der Dauer von betonten und unbetonten Vokalen und eine häufigere Silbentilgung gemessen, was die Autorin als Tendenz zum akzentzählenden Rhythmus interpretiert. Spanisch dagegen zählt sie aufgrund einer geringen Standardabweichung in der durchschnittlichen Vokaldauer und wenigen Tilgungen zu den Sprachen mit silbenzählendem Rhythmus. Das Sprechtempo (es werden phonologisch zugrunde liegende Silben gezählt) liegt in den brasilianischen Aufnahmen mit AGs zwischen 5,45 und 7,10 Silben/s (Mittelwert: 6,57) deutlich unter dem für das Spanische gemessene Tempo (zwischen 7,35 und 9,10 Silben/s, Mittelwert: 7,82). Rebollo Couto sucht in Anlehnung an Quilis (1999) Erklärungen für rhythmische Unterschiede zwischen BP und Spanisch in der Ausgangsposition für die Artikulation der Sprachlaute (damit ist wohl die Lage der Zunge in Ruheposition gemeint), die im Spanischen weiter vorne liegt und damit im Vokalismus Stabilität gewähren soll. In Verbindung hiermit wird die Häufigkeit vorne im Mundraum gebildeter Konsonanten erwähnt. Im BP dagegen seien eine hintere Ausgangsposition bei der Artikulation und eine abnehmende artikulatorische Spannung sowie Diphthongierung und Vokalreduktion zu beobachten. Ob die Ursache für Vokalreduktionen tatsächlich mit der Ausgangsposition für die Artikulation in Zusammenhang gebracht werden kann, ist jedoch fragwürdig.

Cuenca Villarín (2000) vergleicht eine spanische mit einer englischen Lesung des Textes „Nordwind und Sonne“ je eines Muttersprachlers sowie ein informelles Interview.⁶² Das

⁶¹ Eigene Berechnungen nach Angaben in Slembek (1993) in Silben/min bzw. Wörtern/min pro Aufnahme.

⁶² Die angegebenen Werte erfolgen aus eigener Berechnung, Cuenca Villarín gibt Gesamtwerte an.

Spanische weist mit seinen einfacheren Silben eine höhere Silbengeschwindigkeit (SG) auf: $5,65 > 3,64$ Silben/s (vorgelesener Text) und $4,56 > 4,00$ Silben/s (Interview). Außerdem werden metrische Füße (Betonungen) gezählt, wobei sich im Spanischen ein höherer Durchschnittswert für den gelesenen Text ergibt ($2,31 > 1,63$ metrische Füße/s), im spontanen Interview jedoch das Englische einen höheren Wert aufweist ($1,98 > 1,86$ metrische Füße/s). Auch „grupos fónicos“ werden gezählt, wobei nicht definiert wird, ob damit artikulatorische Phrasen oder intonatorische Einheiten gemeint sind. Dabei wird für den vorgelesenen Text im Spanischen eine etwas höhere Geschwindigkeit als im Englischen gemessen ($27 > 23,4$ / min), in dem Gespräch dagegen eine etwas niedrigere ($25,8 < 29,4$ / min).

Dellwo/Wagner (2003) untersuchen die Einteilung in Rhythmuskategorien und stellen einen Unterschied in der SG (Silben/s) zwischen den Sprachen Englisch, Deutsch und Französisch fest, wobei Französisch am schnellsten gesprochen wird und Deutsch etwas langsamer als Englisch.

In Wood (1973a) wird ein Zusammenhang von Silbenstruktur und Silbengeschwindigkeit verdeutlicht. Vier Sprecher (Muttersprachler des amerikanischen und britischen Englisch, Swahili und Westgrönländisch) äußern zwischen 30 und 200 Phrasen. In allen Sprachen ist eine negative Korrelation zwischen der Anzahl der Laute/Silbe und der Silbengeschwindigkeit festzustellen ($r = -0,73$ bis $-0,93$). Da im Westgrönländischen nur durchschnittliche Silbenkomplexitäten von 1,24 bis 1,8 Lauten/Silbe und im Swahili von 0,9 bis 1,5 Lauten/Silbe vorliegen, weisen diese Sprachen höhere Geschwindigkeiten von vier bis acht Silben/s auf, wohingegen im amerikanischen und britischen Englisch Geschwindigkeiten zwischen zwei und acht Silben/s bei mittleren Silbenkomplexitäten von 1,3 bis 2,34 Lauten/Silbe gemessen werden. Für eine Verallgemeinerung der Ergebnisse ist das untersuchte Material nicht umfangreich genug, eine Tendenz ist aber deutlich erkennbar. Wood führt außerdem Ergebnisse aus Sigurd (1971) an, nach denen Versuchsteilnehmer Logatome bei steigender Silbenkomplexität mit abnehmender Wiederholungsrate produzieren. „Ba, spa, stra“ etc. ergeben dabei jeweils abnehmende Geschwindigkeiten von $6,7 > 5,0 > 3,6$ Silben/s.

Aus den erwähnten Arbeiten kann keine Prognose für eine Abhängigkeit des Sprechtempos vom phonologischen Typ abgeleitet werden. Werden in einigen Studien traditionellerweise als „akzentzählend“ bezeichnete Sprachen mit niedrigerem Tempo in Verbindung gebracht, können in anderen Arbeiten gegenteilige Ergebnisse festgehalten werden. Einige Arbeiten vergleichen den Parameter Silben/s und weisen dabei auf einen Unterschied zwischen

Sprachen mit einfachen und solchen mit komplexen Silben hin, der ohnehin auf der Hand liegt. Aufgrund von Silben- und Lautgeschwindigkeit alleine können Äußerungen grundsätzlich nicht als schnell oder langsam bezeichnet werden. Denkbar ist, dass die Messgrößen im Sprachvergleich recht unterschiedliche Ausprägungen aufweisen. Russo/Barry (2008) berechnen allerdings für ihr Material für zwei Sprachen, die unterschiedliche Silbenstrukturen aufweisen, Korrelationskoeffizienten zwischen Silben- und Lautgeschwindigkeit von $r=0,894$ (für Italienisch) und $0,815$ (für Deutsch). Zwar liegt erwartungsgemäß der Korrelationskoeffizient im Deutschen, das komplexere Silben und damit einen größeren Unterschied zwischen Laut- und Silbenzahl aufweisen kann, niedriger, scheint aber dennoch relativ hoch zu liegen.

Bevor der Sprachrhythmus weiter thematisiert wird, soll der Frage nachgegangen werden, ob der Grad der Koartikulation im Sprachvergleich unterschiedliche Ausmaße annimmt, da zunehmende Überlappung eine von mehreren möglichen Strategien zur Beschleunigung der Artikulation darstellt. Dabei muss berücksichtigt werden, dass ein höherer Grad an Koartikulation zwar eine höhere Lautgeschwindigkeit bewirken *kann*, dies aber nicht notwendigerweise die Folge sein *muss*.

3.3 Koartikulation im Sprachvergleich

Manuel (1999) führt eine Zusammenfassung verschiedener Untersuchungen an, auf der die folgenden Ausführungen weitgehend basieren. Dabei wird untersucht, inwiefern phonologische Kontraste in einem Sprachsystem das Ausmaß der Koartikulation bestimmen.

In Clumeck (1976) wird festgestellt, dass während der Artikulation von Vokalen vor nasalen Konsonanten im Englischen und Schwedischen das Gaumensegel früher gesenkt wird als im Französischen, Amoy-Chinesischen und Hindi. In diesen drei Sprachen besteht eine phonologische Opposition zwischen Oral- und Nasalvokalen, so dass bei Oralvokalen nur eine geringe Vorwegnahme des Merkmals „nasal“ zu erwarten ist. Allerdings wird die phonologische Opposition zwischen oraler und nasaler Artikulation in eben dieser Position aufgehoben, so dass nicht zwingend vom phonologischen Kontrast als Determinante der Koartikulation ausgegangen werden kann. Chafcouloff/Marchal (1999) führen auch eine Untersuchung von Rochet/Rochet (1991) an, in der die stärkere nasale Assimilierung im Englischen im Vergleich zum Französischen mit dem phonemischen Kontrast zwischen Nasal- und Oralvokalen im Französischen erklärt wird. Im Französischen wird dort

allgemein schwächere und v.a. regressive Assimilation, besonders bei geschlossenen Vokalen nachgewiesen.

Öhmans (1966) Ergebnisse werden durch die von Choi/Keating (1991) untermauert, nach denen in slawischen Sprachen mit einem phonologischen Kontrast zwischen palatalen und nicht-palatalen Konsonanten in VCV-Abfolgen nur minimale Koartikulation zwischen Vokalen zu beobachten ist. Demnach kann eine deutlich stärkere Koartikulation der Vokale in vergleichbaren Kontexten im Englischen und Schwedischen damit erklärt werden, dass in Sprachen, in denen ein intervokalischer Konsonant bezüglich der Zungenrückenposition spezifiziert werden muss, weniger Koartikulation über den Konsonanten hinweg stattfinden kann.

Lubker/Gay (1982) erklären im Schwedischen früher einsetzende Lippenrundung damit, dass diese einen phonologischen Kontrast darstellt und somit wahrnehmbar sein muss. Sprecher des Englischen führen die Lippenrundung in Konsonanten vor gerundeten Vokalen deutlich später aus, da die Distinktivität des folgenden Vokals nicht auf dem Spiel steht: Hier kontrastieren gerundete und ungerundete Vokale nicht.⁶³ Auch für Mandarin scheint eine Vorwegnahme des Merkmals „gerundet“ wahrscheinlich. Xu (1989) stützt sich auf perzeptive Daten und findet für [ɕ] frühere Lippenrundung vor [y] als für die Konsonanten [ʃ s] vor [u]. Wichtig ist dabei, dass [ɕ] nur vor [i] oder [y] auftreten kann und somit die Diskrimination der Vokale durch eine zeitliche Ausdehnung des (einzigen) unterscheidenden Merkmals erleichtert wird.

Manuel/Krakow (1984) argumentieren, dass in Sprachen mit einem umfangreichen phonologischen Vokalsystem den einzelnen Vokalen weniger phonetischer Spielraum zur Verfügung steht als in Sprachen mit kleineren Vokalsystemen und daher Letztere einen höheren Grad an Koartikulation zulassen – die perzeptive Abgrenzung der Vokalphoneme wird hier aufgrund geringerer Assimilierung (und damit Veränderungen der Frequenzlage) nicht bedroht. Tatsächlich kann nachgewiesen werden, dass Shona und Swahili mit fünf Vokalphonemen mehr Koartikulation zwischen Vokalen zulassen als Englisch. Das Ergebnis wird als mit dem von Magen (1984), die Japanisch und Englisch untersucht, vergleichbar bezeichnet. Auch Manuel (1987; 1990) stellt fest, dass die Größe des Vokalsystems den Grad der Koartikulation bestimmt: Sotho (sieben Vokalphoneme) weist weniger Beeinflussung von Vokalen durch den Vokal der folgenden Silbe auf als Shona

⁶³ Vgl. Daten in Delattre (1965: 104 ff.), aus denen hervorgeht, dass ein Sprecher des Französischen auch in /du/ früher Lippenrundung produziert als ein Sprecher des amerikanischen Englisch, obwohl hier kein Kontrast mit einem ungerundeten Vokal vorliegt.

und Ndebele (fünf Vokalphoneme). Untersucht werden in den genannten Studien die Werte der ersten beiden Formanten. Als Einwand muss hervorgebracht werden, dass im Deutschen, das ein relativ großes Vokalsystem aufweist, spektrale Überschneidungen zwischen den einzelnen Phonemen zu beobachten sind (z.B. zwischen /u/ und /o/ sowie zwischen /ɪ/ und /e/, auch /i/ und /e/; s. Hakkarainen 1995: 122 und Referenzen darin). Die Abgrenzung der einzelnen Vokalphoneme kann damit nicht uneingeschränkt als beschränkender Faktor von koartikulatorischen Effekten angenommen werden.

Wie Manuel berechtigt zu bedenken gibt, kann Koartikulation vom Sprecher in ganz unterschiedlichen Graden eingesetzt werden, womit sprachspezifische Unterschiede nur unter Vorbehalt angenommen werden dürfen. Hinzu kommt, dass große Unterschiede zwischen Sprechern einer Sprache zu beobachten sind, was wiederum die Aussagekraft von Untersuchungen mit wenigen Sprechern einschränkt und widersprüchliche Ergebnisse erklären kann.

Neben den von Manuel (1999) erwähnten Arbeiten werden Untersuchungen zum Thema z.B. von Kochetov et al. (2007) durchgeführt. Drei Konsonantencluster, / kt kp pt /, werden wortmedial und über die Wortgrenze hinweg fünf Mal von drei Sprechern des Russischen und drei Sprechern des Koreanischen bei normalem und hohem Sprechtempo wiederholt; das Sprachmaterial besteht aus bedeutungsvollen Sätzen. In allen Fällen liegt im Koreanischen ein höherer Grad an Überlappung vor. Koreanisch wird als assimilierende Sprache bezeichnet, in der fakultative totale Ortsangleichung in CC-Sequenzen stattfindet, die im Russischen nicht auftritt. Ähnliche Ergebnisse sind bei Zsiga (2000) zu finden. Hier wird eine stärkere Überlappung im Englischen im Vergleich zum Russischen festgestellt (untersucht werden Verschlusslaute über die Wortgrenze hinweg). Auch eine graduelle Palatalisierung von /s/ vor /j/, die im Englischen in einigen Fällen nachgewiesen werden kann, liegt im Russischen nicht vor. Gibbon et al. (1993) messen Koartikulation in / kl /-Sequenzen in „echten“ Wörtern und bedeutungslosen Logatomen in den Sprachen Katalanisch, Englisch, Französisch, Deutsch, Italienisch und Schwedisch (je drei Sprecher) und stellen den höchsten Grad an Überlappung im Katalanischen, den niedrigsten im Schwedischen fest. Auch Yanagawa (2006) findet sprachspezifische Unterschiede: Unter den Sprachen Englisch, Deutsch, Kantonesisch und Japanisch zeichnet sich das Kantonesische als am stärksten überlappende Sprache aus, im Japanischen ist der Grad am geringsten.

Die erwähnten Untersuchungen liefern vereinzelt Hinweise darauf, dass nicht alle Sprachen den gleichen Grad an Überlappung von Segmenten anstreben. Da aber davon

auszugehen ist, dass dieser von vielen Faktoren abhängt, darunter phonologische (z.B. phonologische Kategorien im Sprachsystem, Akzent) und phonetische (Sprechtempo), aber auch weitere sprachliche (Reihenfolge der Segmente und Position im Wort und in der artikulatorischen Phrase; vgl. Gafos et al., 2010 und Tiede et al., 2007) sowie außersprachliche (Sprecher, Register, Sprechsituation, Alter; vgl. Kühnert/Nolan, 1999), ist es empfehlenswert von Tendenzen zu sprechen, was den Einfluss der Sprache auf den Grad der Überlappung anbelangt. Verallgemeinerungen sind nicht haltbar.

3.4 Sprechtempo und Rhythmustypologien

Nachdem in einigen Untersuchungen festgestellt wurde, dass die Silbendauer, auch wenn Tendenzen zu kompensatorischen Kürzungen vorliegen, mit steigender Anzahl der in der Silbe enthaltenen Segmente zunimmt,⁶⁴ ist von einem direkten Zusammenhang von Silbenkomplexität und Silbengeschwindigkeit auszugehen.⁶⁵ Die Silbenkomplexität wird nicht zuletzt seit Dauer (1983; 1987) als ein Kriterium zur Unterscheidung verschiedener phonologischer Rhythmustypen angenommen. Damit kann die Silbengeschwindigkeit als Begleiterscheinung eines Rhythmustyps bezeichnet werden.

An dieser Stelle werden Konzepte vorgestellt, die sich nicht aus phonologischer Sicht mit dem Thema Sprachrhythmus befassen.

3.4.1 Definition von Rhythmus und weiteren relevanten Begriffen

3.4.1.1 Rhythmus

Im Kontext von Sprache wird runter „Rhythmus“ teilweise eine andere Anordnung von Ereignissen verstanden als in der Musik. McAuley (2010: 166) definiert Rhythmus als „the serial pattern of durations marked by a series of events“ und unterscheidet diesen vom wahrgenommenen Rhythmus, der als „perceived temporal organization of the physical

⁶⁴ Kozhevnikov/Chistovich (1965) für Russisch, Sigurd (1971) für Schwedisch, Wood (1973a) für Englisch, Westgrönländisch und Swahili; indirekt Bell-Berti/Harris (1981) für Englisch, Borzone de Manrique/Signorini (1983) für Spanisch, Kohler (1986b) für Deutsch, Crystal/House (1990) für Englisch, Fant et al. (1991) für Schwedisch.

⁶⁵ Neben der Sprechtempomaßeinheit Silben/s wird auch umgekehrt die durchschnittlich benötigte Zeit pro Silbe berechnet, s. Crystal/House (1990) zur „average syllable duration“ (asd). Die Silbendauer unabhängig von der Silbenkomplexität zu betrachten erscheint deshalb sinnvoll, weil lange Silben auch durch Dehnung entstehen können.

sound pattern (i.e., the series of notes and rests)“ definiert wird. Eine ebenfalls an wahrgenommenem Rhythmus orientierte Definition ist die Thomsons:

„The rhythm of any portion of speech is that property of it in virtue of which its movement is consciously apprehended and reproduced as a unity. Or, to make another attempt: The rhythm of a speech-phrase is that complex of properties in virtue of which its movement is apprehended as a unity susceptible of conscious reproduction.“ (Thomson 1923: 141)

Rhythmus existiert demnach nicht von sich aus, sondern wird durch Hörer konstruiert. Wie bei Fraisse (1985) wird hier die Vorhersehbarkeit von Ereigniszeitpunkten impliziert. Dabei wird nicht explizit auf (gleichmäßige) Abstände zwischen Ereignissen eingegangen, vielmehr wird Rhythmus als reproduzierbare und damit von der Gedächtnisleistung und Wahrnehmbarkeit abhängige Struktur verstanden.⁶⁶

Anders als Lenneberg (1986: 147), der Rhythmus mit Isochronie verbindet, setzt Povel (1985) Rhythmizität mit perzeptiver Spannung gleich, die auf Unsicherheit und Ambiguität beruht. Diese Unsicherheit und Ambiguität gehen wiederum darauf zurück, dass in einer Abfolge von Reizen die innere Uhr, die gleichmäßige Abstände zwischen Stimuli zählt, angepasst werden muss, wenn Abstände zwischen Stimuli ungleich sind. Wahrgenommener Rhythmus ist stärker, je stärker die Anpassung der inneren Uhr an das neue (anisochrone, unvorhergesehene) Muster von Reizen ist und je stärker die innere Uhr das vorangegangene Muster verinnerlicht hat. Rhythmus wird demnach als auf Anisochronie beruhende, nicht vorhersehbare Veränderung zeitlicher Abstände zwischen Reizen oder Prominenzen verstanden.⁶⁷ Povel (1985) testet 22 Psychologiestudenten mit musikalischer Ausbildung und lässt 16 Paare von rhythmisch gegliederten Tönen auf ihre Rhythmizität hin beurteilen. Dabei geben die Versuchsteilnehmer an, welche die rhythmischere Sequenz von zwei Stimuli darstellt. Tatsächlich scheinen einige der präsentierten Paare Povels Theorie zu bestätigen. Allerdings stimmen über 90% der Versuchsteilnehmer nur bei zwei Paaren in ihrer Wahl des rhythmischeren Stimulus überein. Das Paar, bei dem die meisten Teilnehmer in ihrer Beurteilung übereinstimmen, weist folgende Rhythmen auf:

(3.1)

(a) I . I . I . . . I I . . . I . . . (b) I . I . I . . . I . . I . I . . .

(nach Povel 1985: 220)

⁶⁶ Damit zusammenhängende rhythmische Gestalten erkannt werden können, müssen die Abstände zwischen Ereignissen im Bereich von ca. 0,1 bis 2,5 s liegen (s. McAuley, 2010: 172 f.).

⁶⁷ Nach obigen Definitionen müssen diese ungleichmäßigen Muster wiederholt werden und damit doch vorhersehbare Zyklen ergeben, um als rhythmisch empfunden zu werden.

Auf den sprachlichen Kontext übertragen bedeutet das, dass Rhythmus nicht nur anhand *einer* Art wiederkehrender Ereignisse im Sprachsignal manifestiert sein kann, sondern möglicherweise durch einen Wechsel prominenter Ereignisse. Bei empirischen Untersuchungen zum Rhythmus kann problematischerweise nicht davon ausgegangen werden, dass alle Versuchsteilnehmer den Begriff „rhythmisch“ gleich interpretieren. Es ist nicht davon auszugehen, dass das, was als rhythmisch empfunden wird, überhaupt nach rein objektiven Kriterien definiert werden kann.

Die Relativität bzw. Subjektivität des wahrgenommenen Rhythmus bestätigt eine Untersuchung von Iversen et al. (2008). Darin wurde herausgefunden, dass die Muttersprache (oder, so die Autoren, die Kultur, in der Sprecher aufwachsen) bestimmt, wie Stimuli unterschiedlicher Länge perzeptiv in Gruppen gegliedert werden. Amerikanische Versuchsteilnehmer gliedern die Abfolge von langen und kurzen Signaltönen in kurz-lang-Abfolgen, während japanische Versuchsteilnehmer keine eindeutige Bevorzugung zeigen. Iversen et al. beschreiben unterschiedliche Erklärungsansätze, darunter syntaktische, gehen dabei aber nicht auf finale Dehnung ein. Diese scheint im Japanischen geringer auszufallen als im Englischen⁶⁸ und hat möglicherweise einen Einfluss auf die metrische Gliederung unterschiedlich langer Stimuli.

3.4.1.2 Rhythmus und Sprache

Um zu den phonologischen Rhythmustypologien überzuleiten soll hier nur eine Arbeit erwähnt werden. Arvaniti/Ross (2010) lassen trochäisch gegliederte Tonfolgen mit Aufnahmen gesprochener Sprache vergleichen. Dabei werden vier Mal zwei Töne präsentiert, die sich in Dauer und Lautstärke (bzw. Schalldruckpegel) unterscheiden ($260 > 120$ ms und $70 > 66$ dB). Nach einer Sekunde Pause hören die Versuchsteilnehmer (22 Amerikaner, 21 Koreaner und 23 Griechen) einen durch Tiefpassfilterung manipulierten Satz. Dabei handelt es sich um eine der Sprachen Deutsch, Englisch, Griechisch, Italienisch, Koreanisch oder Spanisch; dieser Satz besteht in jeder Sprache aus möglichst „silbenzählend“ beschaffenem Material (überwiegend CV-Silben), aus möglichst „akzentzählend“ beschaffenem Material (viele komplexe Silben) oder aus unkontrolliertem Material. Die Teilnehmer vergleichen die rhythmische Ähnlichkeit des unverständlichen Satzes mit dem zuvor präsentierten Muster der Töne und bewerten diese auf einer Skala

⁶⁸ Verhältnis von finalen zu nicht-finalen Silben: 1,24:1 im Japanischen nach Hoequist (1983b), gegenüber 1,53:1 im Englischen nach Delattre (1966); zitiert in Lee/Seoung (1996). In Hoequist (1983a) fällt die finale Dehnung auch im Spanischen stärker aus (29%) als im Japanischen (21%).

von 1 (ungleich) bis 7 (gleich). Weder die Muttersprache der Teilnehmer noch die phonologische Beschaffenheit der Sätze stellen sich als signifikanter Faktor heraus. Allerdings werden die Sprachen zu unterschiedlichen Graden als dem Rhythmus der präsentierten Töne entsprechend bezeichnet, wenn auch nur geringe Unterschiede auffallen: Englisch wird mit einem durchschnittlichen Wert von etwa 3,2 auf der Skala als am wenigsten rhythmisch bezeichnet, das Italienische mit ca. 3,5 als am rhythmischsten. Die anderen Sprachen liegen zwischen 3,4 und 3,45, wobei sich nur Englisch von den restlichen Sprachen statistisch unterscheidet. Die Ergebnisse erinnern zunächst an Gills Einteilung in trochäische und jambische Sprachen, nach der das Englische zum ersten Typ zu zählen ist. Allerdings unterscheidet es sich signifikant vom Deutschen, das Gil ebenso den sogenannten akzentzählenden, trochäischen Sprachen zuordnet. Eine Unterscheidung der untersuchten Sprachen in silben- und akzentzählende Sprachen ist anhand der Ergebnisse nicht möglich. Die Annahme, dass Muster isochroner Töne, die Prominenzen in Bezug auf Dauer und Intensität aufweisen, in ihrer Rhythmizität mit Sprache vergleichbar seien, ist fragwürdig. Daneben handelt es sich bei dem Sprachmaterial um manipulierte Stimuli, aus denen Informationen entfernt wurden, die möglicherweise zur Wahrnehmung von Sprachrhythmus beitragen. Zudem legen die durchschnittlichen Werte zwischen 3,2 und 3,5 auf einer Skala von 1 bis 7 den Verdacht nahe, dass die Versuchsteilnehmer dazu tendierten, möglichst schwache Urteile zu treffen und Extremwerte zu vermeiden, und damit eventuell die Aufgabe als nicht oder nur schwer erfüllbar empfanden oder die Meinungen schlichtweg zu verschieden sind. Das Beispiel soll vor Augen führen, dass der Vergleich des wahrgenommenen Rhythmus von Sprache mit dem von isochronen Tönen nicht angemessen ist.

3.4.2 Rhythmus im phonetischen Sprachsignal: Rhythmusmesswerte

Verschiedene phonetische Eigenschaften standen im Mittelpunkt der Suche nach dem, was sprachliche Rhythmen voneinander unterscheidet. Nachdem Ansätze vorgelegt wurden, in denen die ursprüngliche Dichotomie der akzent- und silbenzählenden Sprachen eingeschränkt oder gar verworfen wird,⁶⁹ halten dennoch viele neuere Untersuchungen an der klassischen Einteilung fest.⁷⁰ In einer Reihe von Arbeiten werden Intervalle von

⁶⁹ Z.B. Dauer (1987), Nespor (1990), Dufter (2003).

⁷⁰ Z.B. Mok/Lee (2008), Lin/Wang (2007), Cho (2004), Dellwo/Wagner (2003) uvm., allerdings sprechen sich einige Autoren nicht explizit für eine strikte Trennung der Rhythmusklassen in akzent-, silben- und morenzählende Sprachen aus, obwohl sie anhand ihrer Ergebnisse derartige Trennungen zu beweisen versuchen. Nespor et al. (2011) unterteilen die Sprachen der Welt in die akzent-, silben- und morenzählenden Klassen, die sich anhand von phonologischen Kriterien unterscheiden lassen. Diese manifestieren sich

konsonantischen und vokalischen Lauten (aber auch stimmhafter und stimmloser Intervalle, z.B. in Dellwo et al., 2007; Galves et al., 2002) zur Abgrenzung von Rhythmustypen untersucht – basierend auf der Annahme, dass diese akustischen Informationen von Ungeborenen im Mutterleib aufgenommen werden, wenn der Rhythmus der Muttersprache internalisiert wird. Die Rhythmusklassenhypothese nach Ramus et al. (1999) und Nazzi et al. (1998) geht davon aus, dass Neugeborene Sprachen unterscheiden können, wenn diese verschiedenen Rhythmusklassen angehören. Mit unterschiedlichem Erfolg scheinen in einigen Arbeiten die Rhythmusklassen akzent-, silben- und morenzählende Sprachen, aber auch intermediäre Sprachen sowie solche mit gemischtem Rhythmus ausgemacht werden zu können.⁷¹ Problematisch sind dabei Auswahl und Umfang der untersuchten Sprachen sowie des Sprachmaterials und der Sprechsituation. In Studien mit einer größeren Anzahl von Sprachen, die teilweise unklassifiziert sind, konnten keine schlüssigen Ergebnisse erzielt werden (Grabe/Low, 2002). Schließlich ordnen unterschiedliche Messgrößen (s.u.) einige Sprachen verschiedenen Rhythmuskategorien zu (z.B. Koreanisch⁷² und Thai, aber auch ursprünglich weniger umstrittene Sprachen wie das Italienische, s. Russo/Barry, 2008; Arvaniti, 2009; Arvaniti/Ross, 2010). Die Suche nach phonetischen Korrelaten phonologischer Rhythmustypen wird deshalb als gescheitert bezeichnet (Arvaniti, 2009) und die Rhythmusklassenhypothese zurückgewiesen (z.B. von Loukina et al., 2011). Seit der Formulierung der Rhythmusklassenhypothese wurden vielzählige Rhythmusmaßeinheiten („rhythm measures“, RM) entwickelt und auf unterschiedliches Sprachmaterial angewendet. Loukina et al. (2011) vergleichen fünfzehn RMs,⁷³ die sie zur Unterscheidung der Sprachen Englisch, Russisch, Griechisch, Französisch und Chinesisch verwenden:

wiederum phonetisch als %V und ΔC .

⁷¹ Dellwo (2009, 2006), Nolan/Asu (2009), Asu/Nolan (2005), Dankovičová/Dellwo (2007), White/Mattys (2007), Barry/Russo (2008), Barry et al. (2003), Wagner/Dellwo (2004), Ferragne/Pellegrino (2004), Mok/Lee (2008), Ramus et al. (1999) uvm.

⁷² Das Koreanische wurde als morenzählend (Cho, 2004), akzentzählend (Lee, 1982; nach Jang, 2009), silbenzählend (Jang, 2009) und Sprache mit gemischtem Rhythmus (Mok/Lee, 2008; Lee et al., 1994) bezeichnet. Jang (2009) zeigt, dass %V und n-PVI-V (s.u.) in verschiedenen SGs geringe Unterschiede aufweisen und klassifiziert Koreanisch diesen Parametern entsprechend, wobei %V mit 54 deutlich über den Werten für silbenzählende Sprachen liegt (48 bzw. 45) und n-PVI-V mit einem Wert von 61 zwischen silben- (36 bzw. 50) und akzentzählenden Sprachen (73 bzw. 82) liegt. Dennoch bezeichnet er Koreanisch als silbenzählend, wenn auch weniger vom akzentzählenden Ideal entfernt als Spanisch und Französisch. Die Referenzwerte bezieht er aus White/Mattys (2007).

⁷³ Weitere RMs wurden für Berechnungen mit stimmhaften und stimmlosen Intervallen im Sprachsignal entwickelt, außerdem wird noch ein RM von Wiget et al. (2010) verwendet (s.u.); Liss et al. (2009) verwenden VC-Sequenzen; Asu/Nolan (2009) auch metrische Füße.

Tabelle 3.1: Übersicht über RMs nach Loukina et al. (2011)

RM	Erklärung	Eingeführt von
%V	Anteil vokalischer Intervalle	Ramus et al. (1999)
Vdur/Cdur	Verhältnis der Dauer vokalischer zu der konsonantischer Intervalle	Barry et al. (2003)
ΔV	Standardabweichung der Dauer vokalischer Intervalle	Ramus et al. (1999)
Varco ΔV	DeltaV / durchschnittliche Dauer vokalischer Intervalle	Dellwo (2006)
VnPVI	normalisierter Variabilitätsindex der Dauer aufeinanderfolgender vokalischer Intervalle	Grabe/Low (2002)
med VnPVI	mit Median berechneter VnPVI	Ferragne/Pellegrino (2004)
ΔC	Standardabweichung der Dauer konsonantischer Intervalle	Ramus et al. (1999)
Varco ΔC	DeltaC / durchschnittliche Dauer vokalischer Intervalle	Dellwo (2006)
CrPVI	Variabilitätsindex der Dauer aufeinanderfolgender konsonantischer Intervalle	Grabe/Low (2002)
CnPVI	normalisierter PVI für die Dauer konsonantische Intervalle	Grabe/Low (2002)
med CrPVI	mit arithmetischem Mittelwert berechneter CrPVI	Ferragne/Pellegrino (2004)
PVI-CV	PVI artikulatorischer Silben	Barry et al. (2003)
VI	Variabilitätsindex der Silbendauer	Deterding (2001)
YARD	Variabilität der Silbendauer	Wagner/Dellwo (2004)
nCVPVI	normalisierter PVI-CV	Asu/Nolan (2005)

Mit Hilfe der RMs werden Klassifikationalgorithmen erstellt, die, auf den Werten der RMs der einzelnen Sprachen basierend, kürzere Ausschnitte einer der untersuchten Sprachen zuordnen sollen. Das untersuchte Korpus ist umfangreich (insgesamt 2300 Texte à durchschnittlich 217 Silben, neun bis zwölf Sprecher pro Sprache) und wurde automatisch segmentiert. Da die Varianzen der gemessenen zeitlichen Ausdehnungen unterschiedlicher Abschnitte (z.B. vokalischer Laute) im Sprachsignal innerhalb einer Sprache enorm sind, schließen die Autoren eine korrekte Identifizierung der Einzelsprachen anhand kürzerer Aufnahmen aus. Es kann kein Klassifikator gefunden werden, der alle fünf Sprachen korrekt identifiziert. Kein einziger RM kann als verlässliches Maß zur eindeutigen Unterscheidung aller Sprachenpaare herangezogen werden, was wie folgt kommentiert wird: „[O]ur results imply that a search for the 'best' RMs for separating all languages is fruitless“ (S. 3265). Ohne signifikante Unterschiede können acht unterschiedliche RMs alleine zu Rate gezogen werden, um einzelne Sprachenpaare voneinander abzugrenzen (%V, Vdur/Cdur, Varco ΔV , nPVI, medVnPVI, VI, YARD und nCVPVI), eine Kombination aus drei RMs stellt sich als ausreichend heraus, um optimale Genauigkeit bei der Unterscheidung aller Sprachen voneinander zu erreichen. Die Einteilung von Sprachen in die herkömmlichen Rhythmuskategorien wird von den Autoren zurückgewiesen:

„[L]anguages from the same rhythm class should always show greater confusion than languages from different classes. We found no such constant pattern of confusion within our five languages. Different combinations of measures produced different patterns of confusion between particular languages or particular subgroups of languages.“ (S. 3266)

Wie Arvaniti (2009), die zeigt, dass Englisch, Spanisch und Griechisch bei geschickt konstruiertem Sprachmaterial fast identische Plätze im $\Delta C/\%V$ -Koordinatensystem einnehmen (s. auch Russo/Barry, 2008 mit ähnlichen Effekten), stellen Wiget et al. (2010) fest, dass RMs, die sich in White/Mattys (2007) als besonders verlässlich zur Unterscheidung von Sprachen (nicht jedoch Rhythmusklassen⁷⁴) herausstellten, in hohem Grade abhängig von der phonetisch-phonologischen Beschaffenheit des untersuchten Sprachmaterials sind. Sowohl die Beschaffenheit eines Satzes als auch der Idiolekt des Sprechers stellen sich als signifikante Faktoren der RMs heraus, so dass festgehalten werden muss, dass die innersprachliche Variabilität der mit den RMs untersuchten Eigenschaften zu groß ist, um Rückschlüsse auf Sprachen zuzulassen. Hierdurch werden v.a. die Ergebnisse von Untersuchungen, die sich auf Korpora mit geringem Sprachmaterial und/oder wenigen Sprechern pro Sprache stützen in Frage gestellt (z.B. Grabe/Low, 2002). Es wird ein weiterer RM vorgeschlagen, der das Ausmaß an Prominenzunterschieden zu differenzieren vermag:

„[T]he Contrast Regularity Index (CRI), a measure of the regularity of sentence stress patterns. We calculated CRI as $1 - (c/t)$, where c stands for the number of syllable transitions in which the stress remains constant (a weak syllable followed by a weak syllable or a strong syllable followed by another strong syllable), and t for the total number of syllable transitions. The lower the number of syllable-to-syllable transitions where stress is unaltered, the higher the CRI, with a CRI of 1 indicating a strict alternation of strong and weak syllables.“ (Wiget et al. 2010: 19)

Es wird nicht ganz klar, was neben der zeitlichen Ausdehnung von Silben unter der phonetischen Umsetzung von „stress“ verstanden wird. Mindestens Unterschiede in der Intensität und der Grundfrequenz können ebenso betonte Silben markieren und sollten nicht außer Acht gelassen werden (s.u.). Außerdem ist ein binärer Kontrast zwischen betonten und unbetonten Silben äußerst fragwürdig, da in einigen Sprachen unterschiedlich starke Nebenbetonungen nachgewiesen werden können.

Barry/Koreman (2009) untersuchen verschiedene Aspekte von metrischem Rhythmus und vergleichen neben der Dauer Grundfrequenz, Intensität und Vokalreduktionen. In einem Produktionsexperiment (Sprachen: Bulgarisch, Deutsch, Englisch; Texte: Lyrik, Prosa und

⁷⁴ White/Mattys (2007: 509) stellen für %V fest: „[T]here was a trend towards Spanish having a significantly higher score than French.“

wiederholte Logatome) wird deutlich, dass das Bulgarische nur geringe finale Dehnung aufweist und im Englischen deutlich höhere PVIs („*Pairwise Variability Index*“, s.o.) für Silben gemessen werden. In allen Sprachen nimmt dieser Wert für Sätze mit kontrastivem Satzakzent gegenüber Sätzen ohne besonderen Fokus („broad focus“) zu. Die Nachahmung der rhythmischen Struktur aus vorgelesenen Gedichten in Sprachmaterial, das aus gleichen, einfachen Silben besteht („dada“-Wiederholungen) weist hohe PVIs für Silben auf (wobei hier auch geringe Unterschiede zwischen den Sprachen bestehen: Im Deutschen fällt die Variabilität etwas geringer aus). Damit steht fest, dass PVIs aufgrund der in der zeitlichen Organisation manifesten Umsetzung von Akzent auch Akzentverhältnisse widerspiegeln und nicht nur temporelle Unterschiede in Silbenstrukturen erfassen, die auf die Anzahl der Laute in der Silbe zurückzuführen sind – zumindest in den untersuchten Sprachen. Daneben wird v.a. die Grundfrequenz zur Akzentuierung verwendet. In einem Wahrnehmungsexperiment zeigen Barry und Koreman, dass die Grundfrequenz einen starken Beitrag zum wahrgenommenen Rhythmus leistet und für bulgarische Versuchsteilnehmer eine größere Bedeutung zu haben scheint als für deutsche oder englische Hörer, was auch auf Intensität und Vokalqualität zutrifft (vgl. Cummins, 2009a und b, im folgenden Abschnitt). Auch hier bestehen große Unterschiede zwischen Versuchsteilnehmern. In einem Satz mit trochäischem Akzentmuster, in dem sich prominente Silben bezüglich der vier untersuchten Parameter unterscheiden, wurden einzelne dieser Parameter neutralisiert, d.h. dass im ursprünglichen Satz aufeinanderfolgende Silben anhand der Dauer (352 vs. 164 ms), Grundfrequenz (H/L-Sequenzen mit Deklination und finalem *Downstep*), Intensität (5 dB Unterschied) und Vokalqualität (/a/ vs. /ə/) differenziert werden können. In einem weiteren Satz werden dann die prominenten Silben durch weniger konturbildende Parameter hervorgehoben (z.B. nur Dauer, Frequenz und Intensität). Unterschiedliche Kombinationen der Parameter ergeben 16 Sätze, einschließlich eines Satzes, in dem alle Silben gleich monoton beschaffen sind. Satz 1 (alle vier Parameter werden zur Hervorhebung akzentuierter Silben verwendet) und Satz 16 (alle Silben sind gleich) werden dann mit den anderen Sätzen paarweise kombiniert, als Stimuli präsentiert, und die Versuchsteilnehmer bulgarischer, deutscher und englischer Herkunft werden angewiesen, anzugeben, welche der beiden Silbenfolgen sie als rhythmischer wahrnehmen und wie stark der Unterschied zwischen den beiden Aufnahmen ausgeprägt ist. Während sich für Deutsche und Engländer die Bedeutung der Dauer zur Wahrnehmung des Rhythmus als größer herausstellt, ist für Bulgaren die Grundfrequenz ebenso wichtig. Vokalqualität und, weniger deutlich,

Intensität, scheinen für bulgarische Teilnehmer auch stärker von Bedeutung zu sein als für Muttersprachler des Deutschen oder Englischen.

Bezeichnet Auer (2001: 1398) noch die zentrale Stellung des Wortakzents für die Isochronieforschung als „ein Konstrukt europäischen (oder sogar eurozentrischen) linguistischen Denkens“ (Klammern im Original), so stellen wenige Jahre später immer mehr Arbeiten die Suche nach phonetischen Korrelaten phonologischer Rhythmustypen mit ihrer überwiegend an temporellen Eigenschaften orientierten Ausrichtung als aussichtslos dar. Dufter (2003: 165) schlägt eine auf phonologischen Überlegungen beruhende Unterscheidung von vier rhythmischen Prototypen vor. Diese Typen verwenden die folgenden rhythmisierenden Mittel, und zwar nicht zu euphonischen Zwecken, sondern um eine optimale Kodierungsleistung zu erreichen:

(3.3)

1. Distinktive Dauer in der Wortphonologie:
Morenbasierter Rhythmus, z.B. Japanisch
2. Distinktive Dauer in der Satzphonologie:
Phrasenbasierter Rhythmus, z.B. Französisch
3. Distinktive Prominenz in Wort- und/oder Satzphonologie:
Prominenzbasierter Rhythmus, z.B. Englisch
4. Keine Kontraste in der rhythmischen Kontur:
Alternierender Rhythmus, z.B. Italienisch

Die rhythmisierenden prosodischen Eigenschaften, die dieser Einteilung zu Grunde liegen, sind Dauer und Prominenz. Dabei wird betont, dass auch in den Sprachen, deren Rhythmus nicht auf Prominenz basiert, keine Isochronie vorausgesetzt (oder angestrebt) wird. In den auf Dauer basierenden Rhythmustypen wird die rhythmische Kontur durch Quantitätsoppositionen festgelegt, was an Hoequists (1983b) Einteilung in komprimierende, kontrollierende und weder komprimierende noch kontrollierende Sprachen erinnert. Darin wird das Japanische als kontrollierende Sprache bezeichnet, da die Dauer einzelner Segmente phonologisch durch die Anzahl der Mora bestimmt ist.⁷⁵ Für den wahrgenommenen Sprachrhythmus würde nach Dufter nahe liegen, dass im Japanischen Dauerverhältnisse für die rhythmische Konturbildung ausschlaggebend sind, was in Anbetracht der Ergebnisse von Iversen et al. (2008; s.o.) jedoch eher unwahrscheinlich ist. Wahrscheinlicher ist, auch unter Berücksichtigung der Ergebnisse von Barry/Koreman

⁷⁵ Auch im Japanischen kann Dehnung unter Akzent (hoher *Pitch*) nachgewiesen werden, die jedoch deutlich geringer ausfällt als im Englischen oder Spanischen, s. Hoequist (1983b). In Anbetracht der Tatsache, dass auf Konsonanten endende schwere Silben bezüglich der Dauer von auf Vokalen endenden abweichen (Hoequist, 1983a), ist eine auf Moren basierende Isochronie unwahrscheinlich.

(2009), dass es Sprachen gibt, in denen nicht (nur) die zeitliche Struktur, sondern mindestens eine der drei anderen prominenzrelevanten Eigenschaften hauptsächlich an der Gestaltung des wahrgenommenen Sprachrhythmus beteiligt ist; dass Prominenz also nicht von phonetischer Länge (im Gegensatz zu phonologischer Länge bei auf Dauer basierenden Rhythmustypen) abhängt, sondern mit anderen Mitteln und möglicherweise einer Kombination aus diesen Mitteln phonetisch manifestiert wird.

Die Vorstellung von Sprachrhythmus nach Auer/Couper-Kuhlen (1994) bezieht sich auch auf im Sprachsignal vorhandene phonetische Ereignisse, die an sich kein rhythmisches Muster darstellen, sondern bei der Verarbeitung durch den Hörer Rhythmisität erlangen. Die Autoren verstehen Rhythmus als notwendigen Beitrag zur Diskursstrukturierung und stellen damit die kommunikative Funktion in den Mittelpunkt. Dabei steht auch die zeitliche Organisation der Sprache im Fokus. Es wird erklärt, dass phonetisch beobachtbare wiederkehrende Ereignisse automatisch rhythmisch gegliedert erscheinen: Hörer suchen Rhythmus, auch wenn er signalphonetisch nicht nachweisbar ist, was bedeutet, dass für die Wahrnehmung von Isochronie nicht die akustische Erkennung von Prominenz vorausgesetzt wird. Im Gegenteil kann auch die Erwartung von Isochronie die Wahrnehmung von Prominenzen auslösen. In diesem Ansatz werden vier Rhythmustypen vorgestellt: Die Abfolge von einfachen, isochronen Reizen; abwechselnd erfolgende starke und schwache Elemente (in gleichen wahrgenommenen Abständen); die Abfolge von starken und schwachen Elementen, bei denen immer die gleiche Zahl an schwachen Elementen zwischen zwei starken Elementen auftritt; und die Abfolge von unterschiedlich langen Gruppen, in denen jeweils ein Element als stark bezeichnet wird, jedoch keine wahrgenommene Isochronie vorliegt. Schematisch werden diese Rhythmen wie in (3.4) veranschaulicht (‘/’ steht für gleiche Abstände, ‘s’ für starke und ‘w’ für schwache Elemente):

(3.4)

$$\begin{array}{cccc}
 /x_i & /x_j & /x_k & /... \\
 /s_i w_{i,l} \dots w_{i,p} & /s_j w_{j,l} \dots w_{j,q} & /s_k w_{k,l} \dots w_{k,r} & /... \\
 /s_i w_{i,l} \dots w_{i,p} & /s_j w_{j,l} \dots w_{j,p} & /s_k w_{k,l} \dots w_{k,p} & /... \\
 s_i w_{i,l} \dots w_{i,n} & s_j w_{j,l} \dots w_{j,n} & s_k w_{k,l} \dots w_{k,n} &
 \end{array}$$

(nach Auer/Couper-Kuhlen, 1994: 86f.)

Sequenzen ohne wahrgenommene Isochronie werden dabei auch als rhythmisch gegliedert bezeichnet. Prominenz und wahrgenommene Isochronie können demnach unabhängig voneinander eigene rhythmische Muster bilden. Auer et al. (1999) veranschaulichen, wie etablierte Rhythmen von Gesprächspartnern übernommen werden, die ihren Redebeitrag

zu bestimmten vom vorgegebenen Rhythmus abhängigen Zeitpunkten beginnen. So entsteht durch die „Schläge“ (Takte) im Sprachsignal zwischen den Gesprächsteilnehmern ein dynamischer interaktionaler Rhythmus. Sprache wird in diesem Ansatz eher als Handeln denn als linguistisches System betrachtet, die Fähigkeit zur Synchronisierung wird vorausgesetzt.

3.4.3 *Entrainment*

Neuere Ansätze aus den Bereichen Sprache, Musik und Biologie verstehen Rhythmus als eine physische und kognitive Befähigung, Bewegungen mit einem wahrgenommenen Takt zu synchronisieren. So betont Cummins (2009a: 17), dass Rhythmus nicht im phonetischen Sprachsignal zu finden ist, nicht einmal in der Sprache selbst, sondern als „*affordance of the entrainment of movement*“ zu verstehen ist. Unter „*affordance*“ versteht Cummins die Beziehung zwischen den Fertigkeiten, die für einen Organismus möglich und lernbar sind, und der Umwelt dieses Organismus. Das Umfeld determiniert, welche Fertigkeiten ein Organismus entwickeln kann. „*Entrainment*“ wird definiert als der Prozess, bei dem sich zwei voneinander unabhängige Schwingungen gegenseitig beeinflussen und eine aufeinander abgestimmte Schwingung entwickeln. Zusammengefasst wird Rhythmus als die vom Umfeld bestimmte Anlage zur Synchronisierung von Bewegung bezeichnet. Für Sprache bedeutet dies, dass sie, sofern sie als rhythmisch bezeichnet wird, für die menschliche Wahrnehmung relevante Eigenschaften aufweist, die den Menschen zur (gleichzeitigen) Produktion befähigen.

In Anwendung auf gesprochene Sprache liegen Hinweise auf persönliche Präferenzen bestimmter Takte in *cycling*-Aufgaben vor. Cummins/Port (1998) lassen Sprecher kurze Phrasen des Typs „*bég for a dime*“ (Akzente geben Betonungen an) zu Signaltönen sprechen. Neben erheblichen Unterschieden zwischen den acht Teilnehmern wird festgestellt, dass die betonten Silben nach der Präsentation der Stimulustöne nicht mit dem vorgegebenen Rhythmus synchronisiert bleiben. Es wird von drei möglichen „Attraktoren“ gesprochen, die Punkte im rhythmischen Zyklus darstellen und betonte Silben „anziehen“. Der erste Attraktor kreiert einen 3/4-Takt, in dem „*bég*“ auf den ersten, „*dime*“ auf den zweiten und eine Pause auf den dritten Schlag fallen. Im zweiten Fall liegt ein 2/4-Takt zugrunde, in dem „*bég*“ und „*dime*“ jeweils auf einen Schlag fallen. Der dritte Attraktor verursacht einen 3/4-Takt, in dem eine weitere Betonung auf „*for*“ eingefügt wird. Bei einer Verringerung der Gesamtzeit eines Zyklus (der Zeit zwischen zwei Signaltönen) verursacht diese Veränderung des Tempos bei manchen Teilnehmern einen Wechsel in einen anderen Takt.

Ein sprachabhängiger Unterschied in derartigen Aufgaben wird von Tajima et al. (1999) nachgewiesen. Muttersprachler des Arabischen, Englischen und Japanischen sprechen kurze Phrasen von bis zu sieben Silben bzw. acht Mora Länge zu Tönen in gleichmäßigen Abständen, die bei aufeinanderfolgenden Durchgängen des Versuchs in kürzeren Abständen präsentiert werden. Die Ergebnisse zeigen, dass genau zwischen zwei Tönen, also nach 0,5 Zyklen, im Japanischen häufig phrasenfinale Silben produziert werden. Im Englischen und Arabischen dagegen werden die letzten betonten Silben der Phrasen an diesem Punkt produziert. Im Englischen werden die betonten Silben außerdem in gleichmäßigeren Abständen produziert als im Arabischen, d.h. die unbetonten Silben zwischen ihnen werden stärker komprimiert. Wird eine rhythmische Gliederung vom Sprecher gefordert, ist es durchaus möglich, dass isochrone Abstände zwischen Betonungen angestrebt werden, wobei dieses Ideal mit unterschiedlichem Erfolg umgesetzt wird. Für das Japanische bedeutet das Ergebnis, dass es als phrasenzählende Sprache bezeichnet werden kann, da keine Kompression innerhalb von Elementen, die in den Phrasen enthalten sind, stattfindet. Tajima et al. (2000) fügen bei der Beschreibung des selben Experiments als Erklärung hinzu, dass im Japanischen Phrasen, die aus einer ungeraden Zahl an Mora bestehen, eine kürzere Pause zur folgenden Wiederholung aufweisen als solche mit einer geraden Anzahl, so dass eine Pause als stummer Schlag den Takt ergänzt. Für die Sprecher des Japanischen wird allerdings von keiner einheitlichen Strategie ausgegangen.

Cummins (2009b) lässt Sprecher einen Text synchron mit einer Aufnahme bzw. einem anderen Sprecher vorlesen. Prinzipiell werden geringere Abweichungen der beiden Sprachsignale erzielt, wenn zwei Sprecher gleichzeitig vorlesen und ihre Sprache miteinander synchronisieren.⁷⁶ Um festzustellen, welche Informationen eine Synchronisierung der eigenen Äußerung mit dem präsentierten Material begünstigen, werden vier Manipulationen getestet: monotonisierte f_0 , Tiefbandfilterung, alleiniges Beibehalten der Hüllkurve und Verstärkung der Intensitätsunterschiede. Es stellt sich heraus, dass große Unterschiede zwischen den Versuchsteilnehmern bestehen. Generell scheint f_0 weniger ausschlaggebend für die Synchronisierung zu sein, da alle untersuchten Sprecher mit monotonisierten Stimuli etwa so präzise synchronisieren wie mit unmanipulierten Aufnahmen. Die Hüllkurve und Intensitätsunterschiede alleine scheinen nicht für eine Synchronisierung ausreichend zu sein; die Ergebnisse für tiefbandgefilterte

⁷⁶ Abweichungen werden anhand eines *Dynamic-Time-Warping*-Algorithmus gemessen. Es werden vier Sprecher untersucht, die über Kopfhörer das eigene Sprachsignal und die Aufnahme, mit der der von ihnen gelesene Text zu synchronisieren ist, hören.

Aufnahmen lassen keine Generalisierungen zu. Die größten Abweichungen werden produziert, wenn keine segmentale Information im präsentierten Material vorhanden ist. In einer Reihe von unterschiedlichen Versuchsbedingungen, in denen Verständlichkeit generell ausgeschlossen ist, kann die Synchronisierung mit (teilweise geringem) Erfolg erzielt werden, so dass davon ausgegangen werden muss, dass Verständlichkeit keine notwendige Voraussetzung für die Synchronisierung ist.⁷⁷ Wird das unverständliche Material ohne segmentale Information z.B. mit dem unveränderten f₀-Verlauf der Originalaufnahme präsentiert, so werden wiederum etwas bessere Ergebnisse erzielt. Schließlich stellt sich auch der Sprecher der Aufnahme, mit der die Versuchsteilnehmer ihre Äußerung synchronisieren sollen, als signifikanter Faktor heraus. Zusammenfassend wird festgehalten, dass Sprecher dazu in der Lage sind, ihre Sprachproduktion mit anderen Sprechern, aber auch mit Aufnahmen zu synchronisieren, wozu verschiedene Informationen des Sprachmaterials mit unterschiedlichem Erfolg genutzt werden können.

3.4.4 Rhythmus und Tempo

Eine Beschleunigung des Sprechtempos scheint sich unterschiedlich auf die segmentale Organisation in Einzelsprachen auszuwirken. Wood (1973b) untersucht, ob bei einer Beschleunigung der Konsonantenteil abnimmt, zunimmt oder gleich bleibt. Kozhevnikov/Chistovich (1966) beobachten, dass der Konsonantenanteil im Russischen zunimmt, also Vokale bei Beschleunigung (stärker) gekürzt und schließlich elidiert werden und damit das Verhältnis von Konsonanten zu Vokalen zugunsten der Konsonanten zunimmt, was auf die artikulatorische Flexibilität der Vokale zurückgeführt wird. Das entspricht Woods Ergebnis für Südschwedisch, allerdings stellt er andere Tendenzen bei Sprechern anderer Sprachen fest: Im Englischen und ägyptischen Arabisch scheint der Anteil größtenteils konstant zu bleiben, in den Sprachen Polnisch, Deutsch und Chinesisch nimmt der Konsonantenanteil von niedrigem zu mittlerem Tempo zunächst zu, dann aber bei einer weiteren Beschleunigung zu hohem Tempo wieder ab. Demnach werden in diesen Sprachen zunächst Vokale, ab einer gewissen, mittleren Geschwindigkeit aber auch Konsonanten reduziert. Offensichtlich werden die Konsonanten erst dann gekürzt und elidiert, wenn eine bestimmte Spannung zwischen den Prinzipien der Verständlichkeit und des möglichst geringen artikulatorischen Aufwandes erreicht wird. Eine gleichmäßige

⁷⁷ Auch in der Fremdsprachendidaktik wird Synchronisierung ohne (vollständige) Verständlichkeit des gehörten und produzierten Materials als Methode eingesetzt.

Kompression von 33% wird von Pickett (1980: 164 f.)⁷⁸ bei einer Beschleunigung der SG von langsamem zu normalem Tempo gemessen. Bei einer weiteren Beschleunigung auf hohes Sprechtempo werden dann jedoch Vokale um 50%, Konsonanten nur um 26% komprimiert. Dabei erfahren kurze Vokale eine stärkere Komprimierung als lange (64% > 43%) und Verschlusslaute eine stärkere als andere Konsonanten (36% > 26%).

Auch auf Silbenebene können ungleichmäßige Kürzungen stattfinden. Lehiste (1970) macht darauf aufmerksam, dass sich betonte und unbetonte Silben unterschiedlich verhalten (vgl. Fletcher, 1987):

„In certain languages, an increase in speech tempo is largely achieved by shortening unstressed syllables, whereas in other languages, the decrease in duration might be proportional over the whole speeded-up utterance. English belongs among the languages of the first type, in which the duration of stressed syllables decreases at a much smaller rate than the duration of unstressed syllables.“ (Lehiste 1970: 38 f.)

Im Schwedischen bleiben dabei Prominenzkontraste gleich wahrnehmbar (Strangert, 1985: 80). Die Unterscheidung von betonten und unbetonten Silben bezüglich ihrer Dauer bleibt über verschiedene Geschwindigkeiten hinweg in ihrem Material (drei Sprecherinnen, je fünf Sätze, drei Tempi) größtenteils konstant.

In dem von Dellwo (2006) untersuchten Korpus (zwölf deutsche, sieben englische und sieben französische Sprecher lesen einen etwa 80 Silben langen Text in fünf unterschiedlichen SGs) stellt sich heraus, dass im Französischen Konsonanten stärker gekürzt werden (auf 45% in der schnellsten verglichen mit der langsamsten Version) als in den anderen beiden Sprachen (auf 57% im Deutschen und 66% im Englischen). Die durchschnittliche Silbendauer ist in allen Tempi im Deutschen am längsten und im Französischen am kürzesten. Ein Vergleich der ΔC -Werte zeigt, dass diese im Deutschen bei zunehmender SG deutlich stärker abnimmt (Differenz: <7) als im Englischen (<2,5) oder Französischen (~2). Dabei ist die Standardabweichung im Deutschen in der langsamsten Version am höchsten und in der schnellsten Version etwa im Bereich des Englischen. Im Französischen liegt die Standardabweichung im Allgemeinen niedriger als in den anderen beiden Sprachen, %V fällt geringfügig höher aus. Eine Unterscheidung silbenzählender von akzentzählenden Sprachen ist anhand der Daten nicht zu erkennen: Nur %V scheint Französisch von Englisch und Deutsch zu unterscheiden, die Verringerung von ΔC trennt jedoch das Deutsche von den anderen beiden Sprachen. Wieder weist das

⁷⁸ Pickett untersucht einen selbst gesprochenen Satz, während das Material bei Wood umfangreicher ist, z.B. ein Auszug von 333 Silben Länge aus einem Roman.

Ergebnis darauf hin, dass Sprecher und Sprachmaterial bestimmen, wie sehr Aufnahmen als „akzentzählend“ bezeichnet werden können. Bestimmte rhythmische Veränderungen als Begleiterscheinung von Veränderungen des Sprechtempo können nicht als Unterscheidungskriterium von Rhythmusklassen herangezogen werden.

Obwohl die für akzentzählende Sprachen typischen Konsonantenhäufungen bei erhöhtem Sprechtempo ab einer gewissen Geschwindigkeit zu Reduktion neigen, scheint eine Strategie zur teilweisen Vermeidung von Konsonantenelision zu beobachten zu sein. Wie für Englisch von Bailey (1980) gezeigt,⁷⁹ stellt Läufer (1995) Resilbifizierungsprozesse bei erhöhtem Tempo im Deutschen fest. Dabei nehmen betonte Silben Konsonanten benachbarter Silben auf, wenn dadurch der Silbenkopf (und damit der Silbenkontakt) verbessert wird, was die Beobachtungen von Wood (1973b) erklären könnte, dass in einigen Fällen die Konsonanten erst bei deutlich erhöhtem Tempo getilgt werden (wenn also selbst Silbenfolgen mit verbessertem Silbenkontakt aufgrund ihrer Komplexität nicht mehr der kanonischen Form entsprechend ohne erhöhten Aufwand artikuliert werden können). Kürzungen werden in unbetonten (leichten) Silben bevorzugt. Für Sprachen mit vielen Konsonantenhäufungen bedeutet das, dass kanonische Formen v.a. bei niedrigem Tempo problemlos realisiert werden können, dieses also bevorzugt wird um die im Sprachsystem ausgenutzte Distinktivität voll umzusetzen. Im Zuge des Sprachwandels aufkommende Konsonantenhäufungen wie mhd. „saf“ > frühnhd. „Saft“ (Szczepaniak, 2007: 249) sind in diesem Sinne als Garanten für ein niedriges Sprechtempo zu interpretieren.

Läufer (1985: 88 ff.) führt Beispiele an, die belegen, dass Resilbifizierungsprozesse bei erhöhtem Tempo die Silbenstruktur verbessern. So tritt in fr. „disloquer“ das /s/ in den Anlaut der zweiten Silbe, was den nicht optimalen Anlaut dieser Silbe stärkt und die vorhergehende Silbe durch Öffnung der Coda verbessert. Gleichzeitig verschiebt sich in fr. „culture“ und de. „Kultur“ die Silbengrenze deshalb nicht, weil dadurch keine Stärkung des zweiten Silbenkopfes erreicht werden kann. Bei erhöhtem Tempo gelten demnach universelle Silbengesetze, die sprachspezifischen Gesetzen, die nur bei normalem Tempo befolgt werden, widersprechen können. Läufer argumentiert, dass diese Verbesserung der Silbe einer Verschlechterung des Kontrastes zwischen Silbennukleus und Silbenrand durch

⁷⁹ /wis.'kan.sm/ wird bei erhöhtem Sprechtempo als /wi.'skan.sm/ resilbifiziert, Bsp. aus Auer (2001: 1393), der auch Fletchers Nachweis (1987) für die qualitative und quantitative Reduktion nicht-akzentuierter Vokale bei höherer SG als Eigenschaft akzentzählender Sprachen anführt. Vgl. auch Stetson (1951: 75), der feststellt, dass wiederholte C₁VC₂-Silben bei Erhöhung der AG zu C₂C₁V-Silben werden, wodurch die schwache Codaposition von C₂ zur starken Anlautposition wird und Schwund vorgebeugt wird.

überproportionale Reduktion der Vokale entgegenwirkt. Gleichzeitig bewirkt in Sprachen mit festem Wortakzent dieser eine Blockierung von Resilbifizierungsprozessen vor der betonten Silbe. In diesen Sprachen wird somit die betonte Silbe im Wort und damit die Gestalt des phonologischen Wortes geschützt, während in anderen Sprachen die Silbenkontur optimiert wird.

Die von Russo/Barry (2008) festgestellte negative Korrelation zwischen Silbengeschwindigkeit und dem Verhältnis von Konsonanten zu Vokalen fällt im Deutschen aufgrund umfassenderer Kürzung der Konsonanten bei erhöhtem Tempo stärker aus als im Italienischen, da das Deutsche mehr Konsonanten aufweist und diese erst bei erhöhtem Tempo reduziert werden, was auch die Zunahme von %V bei erhöhter Lautgeschwindigkeit im Deutschen erklärt. Die Strategien zur Reduktion des Sprachmaterials bei Beschleunigung im segmentellen wie im silbischen Bereich müssen aber, genau wie die Strategien zur Beschleunigung im Allgemeinen, vorerst als sprecherabhängig gedeutet werden, wobei nicht ganz auszuschließen ist, dass Sprachgemeinschaften unterschiedliche Flexibilitäten bezüglich zeitlicher Komprimierung zulassen.

Dellwo (2010) argumentiert, dass gerade aufgrund sprachspezifischer Unterschiede sprachliche Elemente, anhand derer diese Unterschiede besonders deutlich werden, in vergleichenden Untersuchungen zum Tempo herangezogen werden sollten. Die niedrigere Silbengeschwindigkeit des Deutschen und Englischen im Vergleich zum Französischen und Italienischen in einigen von Dellwos Untersuchungen folgt als Korrelar aus der Tatsache, dass die „langsameren“ Sprachen komplexere Silben aufweisen und stellt somit kein neues Ergebnis dar.⁸⁰ Sollte, wie von Dellwo (2010: 65) vermutet, die Lautgeschwindigkeit beim Vergleich der genannten Sprachen tatsächlich keine (oder nur sehr geringe) Unterschiede aufweisen, ist die Darstellung akzentzählender als langsamere Sprachen fragwürdig. Dellwos C-Intervalle können aus mehreren Segmenten bestehen (auch Silbengrenzen überschreitend, s.u.): de. „pflückst“ [pflykst^h] wird bei dieser Betrachtung mit es. „son“ [son] gleichgesetzt. Sowohl „rhythmische“ als auch zeitliche Unterschiede laufen dabei Gefahr, übersehen zu werden. Dellwo verteidigt die Verwendung der C- und V-Intervalle mit der Begründung, dass diese auch von den RMs untersucht werden und der Zusammenhang zwischen Rhythmus und Tempo anhand

⁸⁰ Vgl. Mok (2009: 153), die zusammenfasst: „In addition to consonantal and vocalic durations, this study also demonstrates that syllable durations can potentially be used to distinguish languages with different speech rhythm.“

gleicher Einheiten unternommen werden sollte. Beim Vergleich von fünf intendierten SGs⁸¹ stellt sich heraus, dass Französisch und Italienisch keine signifikanten Unterschiede aufweisen, sich aber bei jedem Tempo von den anderen untersuchten Sprachen (Deutsch, Englisch, Italienisch, Tschechisch) unterscheiden und Englisch sich aufgrund einer geringeren Veränderung der Geschwindigkeit über die fünf Versionen hinweg von allen anderen Sprachen signifikant unterscheidet. Die CV-Geschwindigkeit wird als verlässlicher Indikator der Rhythmusklassse einer Sprache bezeichnet, da statistisch signifikante Unterschiede zwischen Englisch und Deutsch einerseits und Französisch und Italienisch andererseits bestehen. Eine höhere CV-Geschwindigkeit bedeutet, dass einheitlichere (nicht einfachere) C/V-Abfolgen vorkommen, d.h. aufeinanderfolgende Segmente seltener zu einem C- oder V-Intervall zusammengefasst werden können. Dellwo veranschaulicht dies anhand der folgenden Beispiele ('c' und 'v' stehen links für Segmente, 'C' und 'V' rechts für die entsprechenden Intervalle) und berechnet das Verhältnis von Silben zu CV-Sequenzen:

(3.5)

(a) cv cv cv cv cv cv cv cv	8 Silben, 8 C, 8 V, 8:16 = 1:2
(b) cvcc cvc cv cvc ccvc vc cvc cv	8 Silben, 8 C, 8 V, 8:16 = 1:2
(c) cv cv cv cvc cv v vc cv	8 Silben, 6 C, 6 V, 8:12 = 1:1,5
(Dellwo, 2010: 66)	

Bei (3.5a) und (3.5b) handelt es sich um Konstrukte, (3.5c) entspricht einem französischen Satz des Korpus. Das Verhältnis der Silbenzahl zur Anzahl von C- und V-Intervallen fällt im untersuchten Sprachmaterial in den Sprachen Französisch, Englisch und Deutsch bei großer Varianz etwa gleich aus und unterscheidet sich nicht signifikant; im Italienischen ist das Verhältnis bei ebenfalls großer Varianz etwas niedriger. Das Tschechische fällt dadurch auf, dass es ein deutlich weniger ausgewogenes Verhältnis bei auffallend geringerer Varianz aufweist. Entscheidend ist nun, dass bei der Erhebung der CV-Geschwindigkeit „silbenzählende“ Sprachen im Vergleich zu „akzentzählenden“ Sprachen schneller sind. Da eine geringere Anzahl von C- und V-Intervallen auf eine höhere Anzahl von onsetlosen Silben zurückzuführen ist (nur diese erlauben, dass Vokale über die Silbengrenze hinweg zu einem V-Intervall zusammengefasst werden und damit die Anzahl der Intervalle niedrig bleibt), belegt das Ergebnis nicht, dass Silbenstrukturen in „silbenzählenden“ (eigentlich: schnelleren) Sprachen einfacher seien, sondern einheitlicher,

⁸¹ Das Sprachmaterial aus dem zur Untersuchung herangezogenen „BonnTempo-Corpus“ umfasst zwischen 1619 und 5689 Silben pro Sprache und besteht aus von drei bis 15 Sprechern gelesenen Texten pro Sprache.

und „akzentzählende“ Sprachen (besser: Deutsch und Englisch) häufiger keinen konsonantischen Silbenanlaut aufweisen, woraus auch ein hoher Wert für ΔC folgt.⁸² Weil ‘CV VV VC’ zu zwei C- und einem V-Intervall zusammengefasst wird, ‘CV CV CV’ aber als jeweils drei C- und V-Intervalle gezählt wird, folgt aus der zweiten segmentalen Anordnung eine höhere CV-Rate, auch wenn nicht mehr Silben oder Laute/s artikuliert werden. Eine niedrigere CV-Rate erfolgt somit aus höheren Standardabweichungen, was wiederum selbst ein Kriterium zur Unterscheidung der Rhythmusklassen ist. Dellwos französische Beispielsatz (in 3.5c wiedergegeben) ist nicht repräsentativ, wie den Ausführungen indirekt zu entnehmen ist: Das arithmetische Mittel für das Silben-Intervall-Verhältnis liegt im untersuchten Korpus für Französisch bei etwa 1:1,8. Das Tschechische zeigt mit einem Silben-Intervall-Verhältnis von 1:2 ein besonders ausgewogenes Verhältnis, gleichzeitig komplexe aber häufig offene Silben: Nach CV (57%) ist CVC mit 25% der zweithäufigste Silbentyp, gefolgt von CCV mit 8%. CCC-Sequenzen sind mit unter 5% deutlich seltener als im Deutschen (13%) und Englischen (14%; ebd., S. 37). Für die Diskussion um verschiedene präferierte SGs bedeutet dies, dass die CV-Geschwindigkeit, ebenso wie die Silbengeschwindigkeit, keine Vorteile gegenüber anderen Messungen des Tempos bringt und als einzige Darstellungsform des Tempos keine Aussagekraft besitzt. In Bezug auf die Silbenkomplexität wird nur erfasst, wie häufig Silben ohne Konsonanten im Anlaut auftreten. Über die Zuordnung des Tschechischen zu einer Rhythmuskategorie können anhand der Untersuchung keine Aussagen gemacht werden.

Eine Unterscheidung rhythmischer Typen anhand einer einzelnen Dimension des Sprechtempos ist nicht plausibel, sondern folgt aus der Einteilung der Rhythmustypen anhand silbenstruktureller Kriterien. Für die artikulatorische Organisation geben die zusammengefassten Ergebnisse einen Hinweis darauf, dass ein hohes Sprechtempo mit einem begleitenden hohen Konsonantenanteil vermieden wird. Für Sprachen mit einem hohen Konsonantenanteil kann folglich ein niedriges präferiertes Tempo in Aussicht gestellt werden.

⁸² Alleine die Erhöhung von C-Segmenten hätte keine Auswirkung auf die Zahl der Intervalle: In „CVC CVC CVC CVC CVC CVC“ liegen ebenso viele Intervalle vor wie in „CV CV CV CV CV CV“.

3.5 Zusammenfassung

Bisherige Studien belegen nur teilweise Unterschiede im Sprechtempo zwischen Sprachen, wobei keine umfassenden Untersuchungen zu verschiedenen Parametern des Sprechtempos mehrerer Sprachen durchgeführt wurden.

In der segmentellen Organisation der Sprachproduktion können durchaus verschiedene Muster bezüglich der Überlappung von artikulatorischen Gesten, die sprachabhängig zu sein scheinen, festgestellt werden. Allerdings lassen sich aufgrund der Effekte vielzähliger Faktoren keine sicheren Generalisierungen aussprechen. Ein Urteil „Sprache A weist eine höhere Lautgeschwindigkeit auf als Sprache B, weil mehr bzw. stärkere Koartikulation stattfindet“ ist damit trotz einzelner Hinweise unmöglich.

Da Tempo und Rhythmus eng verbunden sind, scheint für die Rhythmusforschung ein relativierender Ansatz aussichtsvoll, in dem verschiedene phonetische Elemente, die auch durch ihr Zusammenwirken konturbildend sein können, berücksichtigt werden müssen. Möglicherweise ist auch das Zusammenspiel verschiedener Sprechtempoparameter von Bedeutung für die wahrgenommene SG. Da Rhythmus erst durch den Hörer entsteht, muss der Gegenstand von einem perzeptiven Standpunkt aus untersucht werden. Des Weiteren sollte berücksichtigt werden, dass persönliche Auslegungen dessen, was als „rhythmisch“ bezeichnet wird, sehr unterschiedlich sein können. Dass sprachliche Stimuli notwendigerweise einen Rhythmus aufweisen, der noch dazu kategorisiert werden kann, darf nicht als gegeben angenommen werden. Ein eindeutiger Zusammenhang des Tempos, d.h. verschiedener Parameter des Sprechtempos mit bestimmten temporell-rhythmischen Eigenschaften wie einer ausgeprägten Präferenz für einfache Silben konnte bislang nicht nachgewiesen werden.

Nachfolgend werden unter Berücksichtigung der phonologischen Systeme und segmenteller Eigenschaften die verschiedenen Parameter des Sprechtempos für sechs Sprachen gemessen.

4 Empirische Studie zum Sprechtempo

4.1 Einleitung

Eine Sprache, in der von intrinsisch kurzen Segmenten und solchen, die mit agileren Artikulationsorganen gebildet werden, häufiger Gebrauch gemacht wird, erfüllt Voraussetzungen für eine relativ hohe Lautgeschwindigkeit. Da deutliche Unterschiede in den Lautinventaren der Sprachen der Welt bekannt sind,⁸³ werden im Folgenden die Phoneminventare und Häufigkeitsverteilungen der Konsonanten sowie phonotaktische Verhältnisse der untersuchten Sprachen unter Berücksichtigung der tatsächlichen Anzahl artikulatorischer Bewegungen und des Höreindrucks näher betrachtet. Die hier angeführten Werte für die Häufigkeit der Konsonanten beruhen auf dem Korpus der empirischen Studie, deren weitere Ergebnisse im Anschluss vorgestellt werden. Laut den in Kapitel 2 erwähnten Studien können mit der Zungenspitze artikulierte Laute schneller artikuliert werden. An dieser Stelle werden die prozentualen Anteile koronal gebildeter Konsonanten angegeben, da sich apikale nicht exakt von laminalen Lauten abgrenzen lassen (vgl. Pompino-Marschall, 2009: 48) und für laminal Laute ebenfalls von potentiell schneller Artikulation aufgrund eines agilen aktiven Artikulationsorgans ausgegangen werden kann. Im Vokalismus wird auf den Anteil nicht-offener oraler Vokale im Phonemsystem eingegangen. Außerdem werden phonologisch relevante silben- und wortoptimierende Erscheinungen dargestellt. Die Ergebnisse dieser phonetisch-phonologisch orientierten Untersuchung werden zunächst für jede Sprache einzeln vorgestellt und anschließend in einem umfassenden Vergleich grafisch dargestellt. Dabei wird auf die Punkte Tempo, Pausenstruktur und temporelle Rhythmuswerte eingegangen.

⁸³ Bezüglich der Größe phonologischer Systeme ist bekannt, dass die Sprache mit dem umfangreichsten Lautinventar (!Xü) 141 Phoneme aufweist (hier werden 34% der Konsonanten vorne gebildet), wohingegen das Inventar des Rotokas 11 Phoneme kennt (66,6% der Konsonanten werden vorne gebildet; s. Pompino-Marschall, 2009: 258 ff.); vgl. auch Häufigkeiten in Delattre (1965: 95 und 97).

4.2 Phonetisch-phonologische Vorstellung der Einzelsprachen: Brasilianisches Portugiesisch, Chinesisch, Deutsch, Englisch, Europäisches Portugiesisch und Spanisch

Brasilianisches Portugiesisch:

Für das in São Paulo gesprochene „Standardbrasilianisch“ kann man, in Anlehnung an Noll (2008) und Cunha/Cintra (2001), von 19 Konsonantenphonemen, 12 Vokalphonemen und zwei Halbvokalen ausgehen:⁸⁴

Tabelle 4.1: Konsonantensystem des BP

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Palatal	Velar
Plosiv	p b		t d				k g
Nasal	m			n		ɲ	
Frikativ		f v		s z	ʃ ʒ		
Vibrant				r			
Getippt				ɾ			
Approximant						(j)	(w)
Lateral				l		ʎ	

Tabelle 4.2: Vokalsystem des BP⁸⁵

i ĩ		u ũ
e ě		o õ
ɛ	ẽ	ɔ
	a	

Nasalvokale sind phonetisch länger, eine Unterscheidung (phonologisch) langer von kurzen Vokalen liegt nicht vor. Die nasale Qualität phonologisch oraler Vokale hängt in der Darstellung von Cristófaró Silva (2007: 121) vom Akzent ab. Reduktionen des

⁸⁴ Die Anmerkungen zur Verwendung der Lautschriftsymbole in diesem Abschnitt gelten ebenso für EP. Der geschlossene Zentralvokal wird hier mit dem Symbol ‘ɐ’ dargestellt, die Halbkonsonanten werden bei den Konsonanten aufgenommen. Die phonetische Realisierung von /r/ schwankt; in São Paulo (Stadt) wird meist [h] bevorzugt.

⁸⁵ Die vereinfachte tabellarische Darstellung der Vokale umfasst die betonten Vokale; dabei werden die Nasalvokale als Phoneme aufgenommen. Obwohl dies umstritten ist (vgl. Cunha/Cintra, 2001: 36), wird hier dieser Standpunkt vertreten, da im BP – im Gegensatz zum EP – nasale Konsonanten nach Nasalvokalen i.d.R. nicht hörbar sind. Minimalpaare wie <vi> [vi] „ich sah“ vs. <vim> [vĩ] „ich kam“, in denen Oral- und Nasalvokal in phonologischer Opposition zueinander stehen, sind zahlreich. Für die Beschreibung der brasilianischen Vokale ist außerdem Cristófaró Silvas (2007: 93) Unterscheidung von nasalierten von nasalen Vokalen von Bedeutung: Bei Ersteren handelt es sich um phonetisch assimilierte Vokale mit nasaler Qualität, bei Letzteren um phonologisch und damit obligatorisch nasale Vokale.

Vokalsystems ergeben sich in unbetonten Silben: Vor der betonten Silbe reduziert sich das Vokalsystem um die halboffenen Vokale / ɛ ɔ /, was dem Vokalsystem in nicht-finalen unbetonten Silben nach der Wortbetonung entspricht. Am Wortende kommen nur die Vokale / i a u / vor. Als Diphthonge werden in Cunha/Cintra (2001) [aj̥ au̥ ej̥ eu̥ iu̥ oj̥ ou̥ ɔj̥ ɔu̥ uj̥

ẽj̥ ẽj̥ õj̥ õj̥] und als Triphthonge [waj̥ wej̥ wiuj̥ wẽuj̥ wõj̥ wõj̥] aufgelistet.⁸⁶ Es werden auch öffnende Diphthonge artikuliert, obwohl diese in den meisten Darstellungen zur Phonetik und/oder Phonologie der Varietäten des Portugiesischen ausgeschlossen werden. Die Unterscheidung von Hiat und Diphthong kann Bedeutungen unterscheiden: pais [pajs] (de. „Eltern“) vs. país [pa'is] (de. „Land“).

33,3% der Vokalphoneme ([i u e o]) zählen zu den kürzeren (nicht nasalen und nicht offenen) Vokalen, und 57,14% der Konsonantenphoneme werden vorne und damit potentiell schneller gebildet. Diese werden häufig gebraucht: Sie stellen 86,84% der artikulierten Konsonanten des Korpus dar.

Tabelle 4.3: Häufigkeit der Konsonanten des BP im gesprochenen Text⁸⁷

Artikulationsbereich	Laute	%	kumulativ
Labial	p b m f v	18,38	18,38
Koronal	t d n l r s ʃ ʒ ɲ ʎ	68,45	86,84
Dorsal	k g ɣ h	13,16	100,00

Den elf von Bisol (1999: 107) dargestellten phonologischen Silbentypen müssen solche mit diphthongiertem Vokalkern hinzugefügt werden:

⁸⁶ Obwohl normalerweise alle vokalischen Bestandteile in nasalen Di- und Triphthongen nasale Qualität aufweisen, wird hier das diakritische Zeichen in diesen Vokalverbindungen aus optischen Gründen nur auf dem offeneren Element angegeben. Außerdem werden entgegen der Darstellungen in Cunha/Cintra (2001) hier die Symbole ‘j̥’ und ‘u̥’ für Halbvokale und ‘j’ und ‘w’ für Halbkonsonanten verwendet.

⁸⁷ Den beweglichen Artikulationsorganen entsprechend wird nur grob zwischen den folgenden Artikulationsbereichen unterschieden: labial (inklusive labiodentalen Lauten), koronal (inklusive interdentalen, dentalen, alveolaren, postalveolaren, retroflexen und teilweise palatalen Lauten) und dorsal (inklusive radikalen Lauten und solchen, die weiter hinten, z.B. glottal gebildet werden).

Tabelle 4.4: Silbenstrukturen im BP mit den entsprechenden Silben in mehrsilbigen Wörtern fett gedruckt

Silbenstruktur	Bsp.	Aussprache	Silbenstruktur	Bsp.	Aussprache
1. V	é	[ɛ]	9. CCVV	gr <u>au</u>	[gr <u>au</u>]
2. VC	ar	[ar]	10. CCVVC	clau stro	[k <u>lau</u> s]
3. CV	cá	[ka]	11. CVVC	m <u>ã</u> os	[m <u>ã</u> ʊs]
4. CVC	lar	[lar]	12. VVC	e <u>i</u> s	[e <u>i</u> s]
5. CCV	trime stre	[trĩ]	13. CVVC	le <u>i</u> s	[le <u>i</u> s]
6. CCVC	tres	[tres]	14. CCVVC	gr <u>au</u> s	[gr <u>au</u> s]
7. VV	aula	[au]	15. CVVV	Ur <u>ugwai</u>	[gw <u>ai</u>]
8. CVV	lei	[lei]	16. CVVVC	sag <u>uões</u>	[gw <u>õ</u> js]

Neben der Palatalisierung der dentalen Verschlusslaute vor palatalen Vokalen und einer selteneren Reduktion im vokalischen Bereich unterscheidet sich das BP durch häufige Vokalepenthese vom EP:

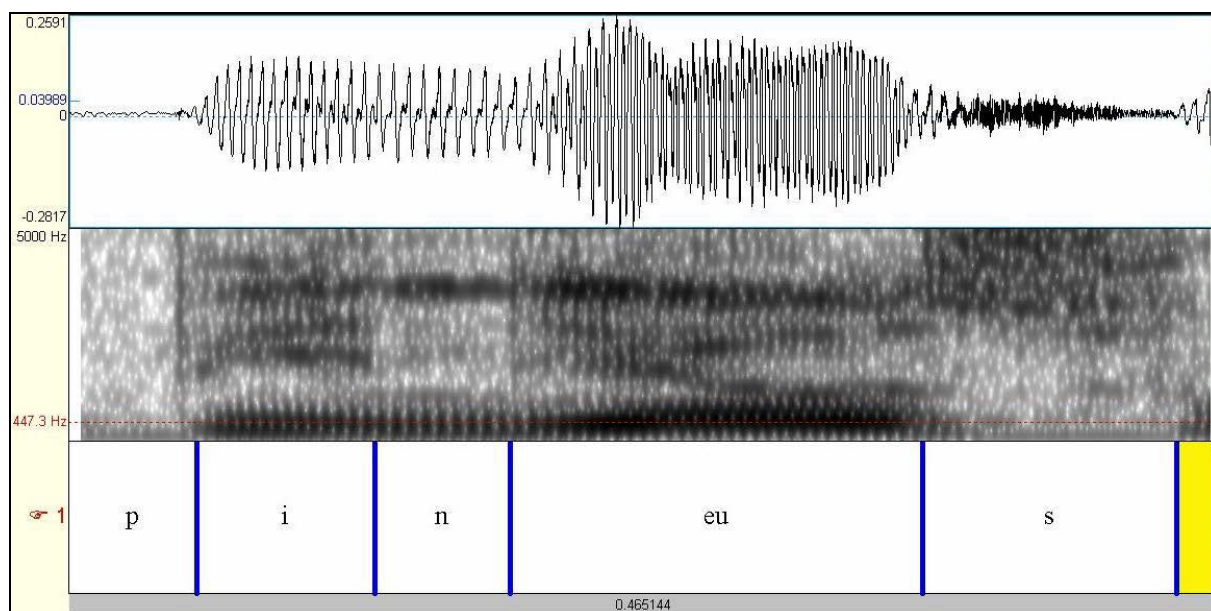


Abbildung 4.1: BP „pneus“ mit Vokalepenthese [pi'neʊs]

Im EP wird „pneus“ (de. „Reifen“) einsilbig [pneuʃ] ausgesprochen. Innerhalb eines Wortes können auch mehrere Silben hinzugefügt werden, wie bei dem Namen der Kaufhauskette „fnac“, der [fi'naki] gesprochen wird. Derartige Beispiele verdeutlichen, dass der Silbe im BP eine bedeutende Rolle als phonologische Kategorie zukommt, obwohl auch wortphonologische Züge wie Reduktion des Vokalismus in Nebensilben und Elision in Endsilben⁸⁸ („noite“ > [noĩtʃ]) vorhanden sind (vgl. Reich, 2004). Mattoso Camara Jr. (1972: 56) beschreibt außerdem silbenoptimierende Resilbifizierungsprozesse

⁸⁸ S. Noll (2008: 57).

für BP und EP. In den Aufnahmen dieser Arbeit weist das BP mit einer mittleren Anzahl von 2,33 Lauten/Silbe und einem Anteil von 30,9% CV-Silben relativ einfache Silben aber eine mittlere Variabilität der segmentalen Struktur derselben auf.

Chinesisch:

Im Chinesischen besteht der phonemische Unterschied zwischen Verschlusslauten, die am gleichen Ort gebildet werden, nicht wie in den romanischen Sprachen in einer Stimmttonopposition oder wie in den germanischen Sprachen in einer Gespannthheitsopposition (s.u.), sondern in der Aspirierung (Deterding/Nolan, 2007). Von den 18 anlautenden Konsonantenphonemen sind nur vier stimmhaft, weshalb eine Tendenz zur Optimierung von Silben angenommen werden kann. Da außerdem silbenfinal nur zwei Phoneme, / n ŋ /, auftreten⁸⁹ und diese häufig einer Abschwächung in ihrer Konsonantenstärke unterzogen werden (fehlender Kontakt der Artikulatoren, s. Halliday, 2006[1992]; häufig auch totaler Schwund bei Nasalisierung des vorausgehenden Vokals), liegt die Vermutung nahe, dass es sich um eine relativ strikte CV-Sprache handelt, in der fast ausschließlich offene Silben mit stimmlosen Silbenköpfen vorkommen. Bei einer genaueren Untersuchung des Phoneminventars und der phonotaktischen Verhältnisse wird jedoch klar, dass dies nicht der Fall ist.

Tabelle 4.5: Konsonantensystem des Chinesischen⁹⁰

	Bilabial	Labiodental	Dental	Alveolar	Palatal	Velar
Plosiv	p p ^h		t t ^h			k k ^h
Nasal	m		n			(ŋ)
Frikativ		f	s	ʃ ʂ		x
Affrikativ ⁹¹			ts ts ^h	tʃ tʃ ^h		
Approximant					(j) (ɥ)	(w)
Lateral			l			

⁸⁹ Der Status von /ɹ/ als silbenschließendem Konsonanten ist aufgrund der approximantischen Qualität und der Distribution umstritten, Duanmu (2007: 40, 212 ff.) analysiert den Laut als vokalisches Suffix [ə].

⁹⁰ In Anlehnung an Hunold (2005) und IPA-Darstellungen; Approximanten in Klammern, da ihr Status als konsonantischer Anlaut fragwürdig ist; der velare Nasal ebenfalls in Klammern da er nicht silbeninitial vorkommt und als Nasalisierung des vorangehenden Vokals realisiert werden kann. Vgl. außerdem Beschreibungen bei Ladefoged/Maddieson (1996: 164) und Lee (1999) der Retroflexe als laminale Sibilanten.

⁹¹ Die Affrikaten werden hier zur Verbesserung der optischen Form ohne Bogen dargestellt.

Tabelle 4.6: Vokalsystem des Chinesischen nach Duanmu (2007: 35 f.)

i y		u
	ə	
	a	

Das Chinesische kennt auch einige Di- und Triphthonge: [aᵛ eᵛ oᵛ ja je wa wo ɥe jo] und [wai̯ wei̯ jau̯ jou̯].⁹² Duanmu (2007) erwähnt Beschreibungen von Wang (1993), in der nur zwei Vokalphoneme gezählt werden und von Pulleyblank (1984), in der das Chinesische über gar keine Vokalphoneme verfügt. Die Allophonie kann im vokalischen Bereich stark differenziert beschrieben werden, so geht man mindestens von den Allophenen [e ɛ o ɔ ə ʏ] für /ə/ und [A⁹³ a æ a ɐ] für /a/ aus, was als Hinweis auf einen hohen Grad an Koartikulation interpretiert werden kann.

Die palatalen Konsonanten [tɕ tɕʰ ɕ] werden teilweise als eigenständige Phoneme, meist als Allophone der dentalen, retroflexen oder velaren Reihe (aus der sie historisch entstanden sind) angesehen, da sie nur in Verbindung mit palatalen Vokalen auftreten (dazu Duanmu 2007: 31 ff.). Unter Berücksichtigung der palatalen Konsonanten kann festgehalten werden, dass sechs Konsonanten aus zwei hier als Segmente gezählten Bestandteilen bestehen (Verschluss und Reibung). Bei drei weiteren Phonemen tritt noch eine Aspiration zu diesen beiden Artikulationsbewegungen hinzu. Im gesprochenen Text werden 76,20% aller Konsonanten vorne artikuliert.

Tabelle 4.7: Häufigkeit chinesischer Konsonanten im gesprochenen Text⁹⁴

Artikulationsbereich	Laute	%	kumulativ
Labial	p p ^h m f	8,56	8,56
Koronal	t t ^h l n ts ts ^h s tɕ tɕ ^h ɕ	67,64	76,20
Dorsal	k k ^h x ŋ	23,80	100,00

Von den phonetischen Vokalen können 46,9% als relativ kurz bezeichnet werden ([i y u e ə ʏ]). Besondere Aufmerksamkeit verdient auch der „apikale“ Vokal [ɿ] (nach heutigen IPA-Konventionen [ɨ]), der von Duanmu als eine Art silbische Verlängerung des vorausgehenden Konsonanten beschrieben wird: [sz] ohne Vokalkern. Dabei bleibt

⁹² Da der initiale Halbkonsonant in Triphthongen meist zum Silbenkopf gerechnet wird, erscheinen in vielen Darstellungen, u.a. der von Duanmu (2007), keine Triphthonge.

⁹³ ‘A’ beschreibt hier einen „central low vowel“, der offener ist als [ɐ] (Duanmu 2007: 38). Bei Ausfall auslautender Nasalkonsonanten können entsprechende Minimalpaare anhand des Vokals unterschieden werden: 单 (/tan/), de. „einzeln“ vs. 当 (/taŋ/), de. „fungieren“, werden als [tæ̃] und [tā̃] realisiert.

⁹⁴ Sowohl aspirierte als auch nicht aspirierte Verschlusslaute werden hier für den entsprechenden Artikulationsbereich aufgelistet, da die Behauchung an der gleichen Stelle produziert wird.

umstritten, ob in 四 (sì, de. „vier“) nicht eine Öffnungsbewegung vorliegt, aufgrund derer der auf den konsonantischen Silbenbeginn folgende Laut zumindest nicht als stimmhaftes Pendant zum initialen Konsonanten bezeichnet werden darf. Phonologisch ist es plausibel, von konsonantischen Silbenkernen auszugehen, da Wörter (die als solche in Wörterbücher zu finden sind) wie 恩 (de. u.a. „ja/ähm“) nur aus einem Konsonanten bestehen – auch wenn das hier genannte Beispiel in Pinyin als „en“ (mit verschiedenen Tönen) transliteriert wird.

Ein phonologischer Vokallängenkontrast wird von Duanmu (2007: 40 f.) ausgeschlossen und als durch die Akzentverhältnisse vorhersehbar bezeichnet. Problematisch ist dabei, dass Minimalpaare gebildet werden können, in denen bei gleicher Betonung die Bedeutung nur anhand der Tonhaftig- bzw. Tonlosigkeit der Silben unterschieden werden kann und tonlose Silben mit einem kurzem Vokal gesprochen werden (dessen Qualität reduziert werden kann); in 到底 (dàodi, de. „eigentlich“) vs. 道地 (dàodi, de. „Unterführung“) kontrastiert eine tonlose mit einer tontragenden Silbe (dieses und weitere Bsp. bei Lin, 2001). Sofern die Abwesenheit eines phonologischen Tons nicht als fünfter Ton anerkannt wird, liegt in diesen Minimalpaaren ein Vokallängenkontrast vor, der dann phonologisch zu bewerten ist. Nur wenn die tonlose Silbe als lexikalisch festgelegt bezeichnet wird, ergibt sich als *sekundäres* Merkmal ein Unterschied in der Vokallänge.⁹⁵ Auch tontragende Silben unterscheiden sich in der Länge. Der vierte (fallende) Ton geht mit kürzerer Dauer einher, der dritte, fallend-steigende Ton dagegen fordert eine längere Silbendauer (vgl. Duanmus Analyse von Silben im dritten Ton als „superheavy“, 2011a: 2757; im Kontext, d.h. nicht isoliert gesprochen, wird der dritte Ton zu einem „halben dritten“, tief-fallenden Ton gekürzt). Außerdem zählt Duanmu etwa ein Drittel aller Silben im gesprochenen Chinesisch zu den tonlosen Silben, die zu den kürzesten gehören (phonologisch betrachtet kommt dieser Ton weitaus seltener vor, in der Nachrichtensprache werden tonale Informationen eher bewahrt als in der Umgangssprache). Es besteht offensichtlich eine Tendenz dazu, kurze Silben zu verwenden, die die Silbengeschwindigkeit erhöhen, was aus der Häufigkeit der Töne in den Nachrichtensendungen hervorgeht:⁹⁶ T4 (33,6%) > T1 (24,9%) > T2 (21,1%) > T3 (13,3%) > T5 (7,0%).

⁹⁵ Chao (1968: 36 f.) erwähnt außerdem Neutralisierung in tonlosen Silben sowie Sonorisierung der stimmlosen Konsonanten, was eine Schwächung der Silbenstruktur und mögliche Stärkung des Wortes bedeuten kann: 篱笆 (líba, de. „Zaun“), [li.pa] > [li.bə], was in Sievers' (1901) Sinn als eine einzige Drucksilbe zählt.

⁹⁶ Duanmu (2011a: 2772) listet diese Häufigkeiten der Töne auf: T4 (33,5) > T2 (25,1) > T1 (23,5) > T3 (17,8) > T5 (0,2).

Darstellungen der chinesischen Silbentypen umfassen meist wenige Strukturen; in Anlehnung an Qiu (1984) listet Hunold (2005: 22) zwölf Silbentypen auf.⁹⁷ Die komplexen Artikulationsbewegungen berücksichtigend ergeben sich folgende Silbentypen:

Tabelle 4.8: Silbentypen des Chinesischen⁹⁸

Silbenstruktur	Zeichen	Aussprache	Silbenstruktur	Zeichen	Aussprache
1. V	一	[i]	10. CVVV	表	[pjaŋ]
2. VC	安	[an]	11. CCV	杂	[tsa]
3. VV	爱	[aɪ]	12. CCVV	走	[tsou]
4. VVC	要	[juŋ]	13. CCVVV	叫	[tɕjaŋ]
5. VVV	用	[jaŋ]	14. CCVC	中	[tʂuŋ]
6. CV	八	[pa]	15. CCVVC	庄	[tʂwaŋ]
7. CVV	得	[tɕɛ]	16. CVCC ^c CC ^c	辛苦	[ɕink ^h wx ^w]
8. CVC	更	[kəŋ]	17. C ^c CC ^c CVV	土地	[t ^h wx ^w tʃi]
9. CVVC	点	[tʃen]	18. C ^c CVVVCC	瞧着	[t ^h ʃjaŋtʂ]

Unter diesen 20 Silbentypen befinden sich komplexe Silben, die aus maximal acht Lauten bestehen⁹⁹, darunter zwei Gleitlaute. Damit kann die Silbe im Mandarin als komplexer bezeichnet werden als es zunächst den Anschein erweckt, was die traditionelle Zuordnung des Chinesischen zu den silben- bzw. morenzählenden Sprachen (s. Hunoldt, 2005; Lin/Wang, 2007) in Frage stellt. Auch die Zuordnung zu den Silbensprachen¹⁰⁰ ist nicht uneingeschränkt vertretbar, da neben der Vermeidung silbenoptimierender Verschiebungen von Silbengrenzen (Duanmu, 2009: 33) eine Reihe phonetischer Reduktionen (segmental, wie Vokalelision, als auch suprasegmental, wie Tonverlust¹⁰¹) innerhalb von phonologischen Wörtern stattfindet. Diese Prozesse schaffen weitere komplexe Silben mit Konsonantenhäufungen in der Silbencoda: In 男子 (nánzi, de. „Mann“) resultiert aufgrund der Vokalelision in der unbetonten Silbe [nants] (vgl. Yen, 1992: 257).¹⁰² Nach Yen (1992:

⁹⁷ Weiter reduziert werden die Silbentypen in Duanmus CVX-Theorie (u.a. 2011b), in der Gleitlaute nach initialem Konsonanten als artikulatorische Geste einer komplexen C-Position gewertet werden ([s^w] wie in 岁, suì, de. „Lebensjahr“, wird als Kopf mit sekundärer konsonantischer Artikulation einer CV-Silbe dargestellt, s. Duanmu 2007: 80), ebenso weitere konsonantische Bestandteile wie z.B. Aspirierung (s.u.). Einwände gegen die CVX-Theorie, die eine universelle Silbenstruktur mit diesem Schema fordert, werden von Marlo (2004) und Gregová (2011) vorgebracht.

⁹⁸ Silbeninitiale öffnende Diphthonge werden gängigen Beschreibungen entsprechend berücksichtigt, auch wenn Anlautbildung nicht ausgeschlossen werden kann (VVV > CVV).

⁹⁹ Wenn Behauchung, wie weiter unten dargestellt, als eigener Laut gezählt wird, ergeben sich fünf weitere Silbentypen: CCCV (擦, [ts^ha]), CCCVV (草, [ts^hau]), CCCVVV (桥, [tɕ^hjaŋ]), CCCVC (唱, [tʂ^haŋ]), CCCVVC (前, [tɕ^hjen]).

¹⁰⁰ In der Terminologie von Auer (2001) stellt in „Silbensprachen“ die phonologische Silbe die Hauptkategorie phonologischer Prozesse dar, während in „Wortsprachen“ das phonologische Wort diese Rolle einnimmt, s. a. Auer (1994).

¹⁰¹ Vgl. Duanmu (2007), der Ton nicht als suprasegmentale Eigenschaft sieht.

¹⁰² Ein weiteres vielzitiertes Beispiel ist 我们 (women, de. „wir“) mit der Aussprache [wom], s. Duanmu

255) treten aus Kontraktionen entstandene Konsonantenhäufungen auch im Anlaut auf, z.B. 时代 (shídài, de. „Epoche“), [ʃtʰaj]. Die Silbentypen 16.-17. (nach Duanmu 2007: 92) und der selbst konstruierte Typ 18. müssen nicht als Bestandteil der Normsprache angesehen werden. Zu erheblichen Zweifeln berechtigt dabei die Tatsache, dass das 1:1-Verhältnis zwischen Morphem und Silbe bei den genannten Silben systematisch durchbrochen wird, obwohl Duanmu (2011a: 2770) hierzu weitere Beispiele auflistet. Auch wenn die zuletzt genannten Beispiele aufgrund ihrer Seltenheit und Konstruiertheit nicht berücksichtigt werden ist offensichtlich, dass es im gesprochenen Chinesisch zu komplexen Aneinanderreihungen konsonantischer Artikulationsbewegungen kommen kann.

Eine weitere Besonderheit des Chinesischen sind die phonotaktischen Beschränkungen. In komplexen Silben können die Artikulationsstellen der einzelnen Segmente nicht weit auseinander liegen. Velare Konsonanten z.B. können nicht mit palatalen Vokalen kombiniert werden. Die Vokale /u/ und /a/ werden nach palatalen Konsonanten zu [y] und [æ] assimiliert. Darüber hinaus kann hinter diesen assimilierten Vokalen ein silbenfinaler Konsonant nur alveolar sein: 卷 (juǎn, de. „Rolle“), [tɕʰæ̃n]. *[tɕʰæ̃ŋ] stellt keine mögliche Silbe des Mandarin dar. Weiterhin kann in Triphthongen vor und nach dem Vokal nicht der gleiche Gleitlaut stehen. Diese Beschränkung ermöglicht, da die Zungenposition nur in eine Richtung verändert wird, einen unbeschwerlichen Bewegungsablauf. Würde die Zunge im Mund vor und zurück bewegt werden ([jaɪ]¹⁰³), wäre damit ein höherer Aufwand verbunden, sofern alle Zielkonfigurationen erreicht werden sollen. Die Beschränkungen haben eine niedrige Anzahl von tatsächlich verwendeten Silben zur Folge (404 Silben ohne Berücksichtigung der Töne laut Duanmu, 2011a: 2763; im Vergleich zu 3219 im Englischen¹⁰⁴). Es ist damit ein Bestreben in der Sprache festzustellen, Artikulationsorte und Öffnungsgrade in erhöhtem Maße einander anzupassen (vgl. hierzu auch die in 3.3. erwähnte Antizipation phonologischer Merkmale). Zwar ist für alle Sprachen generell von Koartikulation beim Sprechen auszugehen, die relativ stark ausgeprägte Neigung zu Assimilationen kann im Chinesischen aber ein erhöhtes Sprechtempo begünstigen, da weniger komplexe Artikulationsbewegungen ausgeführt werden müssen, bei denen zu entfernten Artikulationsstellen und stark abweichenden Öffnungsgraden (im Vergleich zu

(2009: 81).

¹⁰³ 崖, [jaɪ] (de. „Klippe“) wird oft als Ausnahme zu dieser Regel erwähnt, erscheint aber in einigen modernen Wörterbüchern nur noch als [ja].

¹⁰⁴ Duanmu nennt in einer online-Version seines Artikels (Duanmu 2011c) Baayen et al. (1995) als Quelle für diese Zahl. Dazu muss angemerkt werden, dass nach Schiller et al. (1996: 25) auch im Englischen und Niederländischen über 80% aller Äußerungen aus den 500 häufigsten Silben bestehen, die unter fünf Prozent aller in diesen Sprachen vorkommenden Silbentypen darstellen (vgl. Schiller, 2002: 19).

benachbarten Lauten) übergeleitet werden müsste. Duanmu (2007: 60) erklärt damit u.a. die Bedingung „Rhyme-Harmony“ zur Wohlgeformtheit chinesischer Silben, nach der im Silbenreim keine gegensätzlichen Merkmale für Rundung und Artikulationsort kombiniert werden können.

Es bleibt festzuhalten, dass das Chinesische paradox erscheinende Tendenzen aufweist, die einerseits eine höhere AG begünstigen können (offene Silben, Reduktion längerer Töne zu kürzeren, viele vordere Konsonanten), andererseits dieser entgegenwirken (komplexe Konsonantenbewegungen). In den Auswertungen wurden eine durchschnittliche Anzahl von 3,13 Lauten/Silbe und ein CV-Silbenanteil von 17,39% gemessen.

Deutsch:

Für das Lautsystem des Deutschen können mindestens 20 Konsonanten- und mindestens 15 Vokalphoneme angesetzt werden.¹⁰⁵ Sie lauten:

Tabelle 4.9: Konsonantensystem des Deutschen (in Anlehnung an Pompino-Marschall, 2009: 265)

	Bilabial	Labiodental	Alveolar	Postalveolar	Palatal	Velar	Uvular	Glottal
Plosiv	p b		t d			k g		(ʔ)
Nasal	m		n			ŋ		
Frikativ		f v	s z	ʃ ʒ	ç		ʁ	h
Approximant					j			
Lateral			l					

Tabelle 4.10: Vokalsystem des Deutschen (nach Pompino-Marschall, 2009: 266)

i ɪ y ʏ		ʊ u
e ø		o
ɛ œ ɛ:	(ə)	ɔ
	a a:	

Zu den Konsonanten ist hinzuzufügen, dass die Behauchung bei gespannten Verschlusslauten in den Nachrichtenaufnahmen häufig ist und diese im absoluten Auslaut fast ausnahmslos stark aspiriert werden.¹⁰⁶ Dabei kann eine sprecherabhängige Variation

¹⁰⁵ Eine gewisse Uneinigkeit besteht bei der Wertung des Glottisschlages und des Schwas. Affrikaten werden nicht monophonematisch bewertet (Kohler, 1995: 166 f.), neben den drei „echten“ Diphthongen (s.u.) werden hier keine aus der Vokalisierung von /r/ entstehenden Diphthonge aufgelistet, die in Verbindung mit allen Vokalphonemen entstehen können. Bei der Darstellung der Vokale wird die Länge bei gespannten Vokalen als Sekundärmerkmal behandelt. S. hierzu u.a. Antoniadis/Strube (1984), wo auch festgestellt wird, dass in Abhängigkeit von der Zungenlage die Vokaldauer der gespannten Vokale zwischen neun und 19 ms, die der ungespannten Vokale zwischen fünf und acht ms schwankt.

¹⁰⁶ Jessen (1998: 92 ff.) zeigt, dass phonetisch betrachtet die Behauchung eher dazu geeignet ist als die Stimmhaftigkeit, die phonologische Opposition zwischen gespannten und ungespannten Verschlusslauten darzustellen; für Reibelaute sind Dauer und Stimmhaftigkeit geeignet. In der Tabelle werden diese

beobachtet werden, weswegen die finale Behauchung bei der Auflistung der Silbentypen unten nicht berücksichtigt wird.

Als echte Diphthonge gelten [aɪ aʊ ɔɪ], Triphthonge kennt das Deutsche nicht. Von den 20 Konsonantenphonemen werden 55% im vorderen Artikulationsraum gebildet, unter den Vokalen zählen 43,75% ([ɪ ʊ ɛ œ ə]) zu den potentiell kürzeren. Die Häufigkeit vorne gebildeter konsonantischer Laute liegt im Deutschen bei 80,99%:

Tabelle 4.11: Häufigkeit deutscher Konsonanten im gesprochenen Text¹⁰⁷

Artikulationsbereich	Laute	%	kumulativ
Labial	p b m f v	18,51	18,51
Koronal	t d n l s z ʃ ʒ ʧ	62,48	80,99
Dorsal	k g x ŋ ɣ h ʔ	19,00	99,99

Das Deutsche weist eine hohe Zahl an Silbentypen auf. Die einfachste Silbe besteht dabei aus einem Monophthong, in komplexeren Silben können bis zu drei Konsonanten vor und bis zu fünf Konsonanten hinter diesem Silbenkern auftreten. Damit ergeben sich folgende Silbentypen:

Unterscheidungen außer Acht gelassen.

¹⁰⁷ Die hier gemessenen Häufigkeiten weichen teilweise von denen in anderen Untersuchungen ab. Die relative Häufigkeit der konsonantischen Laute lautet absteigend: / n t s d r l m f g b k ʃ v x p h ŋ z ʔ ʒ z /. Zum Vergleich die Reihenfolge nach Kohler (1995: 222): / t n r s l ʔ f g k ʃ b d m p v ŋ x z h ʒ / und nach Delattre (1965: 95): / n t r d s l x m f v g z b k ʃ h p ŋ j /.

Tabelle 4.12: Silbentypen im Deutschen¹⁰⁸

Silbenstruktur	Bsp.	Aussprache	Silbenstruktur	Bsp.	Aussprache
1. V	eh	[e]	20. CVVCC	rauft	[ʁaʊft]
2. VV	au	[aʊ]	21. VCCCC	ernst	[ɛʁnst]
3. VC	an	[an]	22. CCCVC	Spruch	[ʃpʁʊx]
4. VVC	aus	[aʊs]	23. CCCVVC	Streich	[ʃtraɪx]
5. CV	da	[da]	24. CCVCC	brüsk	[bʁʏsk]
6. CVV	rau	[ʁaʊ]	25. CCVVCC	graust	[gʁaʊst]
7. VCC	Abt	[abt]	26. CVCCC	bannst	[banst]
8. VVCC	eins	[aɪns]	27. CVVCCC	raufst	[ʁaʊfst]
9. CCV	dreh	[dʁe]	28. CCVCCC	prompt	[pʁɒmpt]
10. CCVV	drei	[dʁai]	29. CCVVCCC	brauchst	[bʁaʊxst]
11. CVC	dann	[dan]	30. CCCVCC	Pfropf	[pfʁɒp]
12. CVVC	raus	[ʁaʊs]	31. CCCVVCC	streicht	[ʃtʁaɪxt]
13. VCCC	Obst	[ɒbst]	32. CVCCCC	Herbsts	[hɛʁbsts]
14. VVCCC	einst	[aɪnst]	33. CCVCCCC	schrumpft	[ʃʁʊmpft]
15. CCCV	Spre	[ʃpre]	34. CVVCCCC	feilscht	[faɪlʃt]
16. CCCVV	Spreu	[ʃpʁɔ]	35. CCCVCCC	Strumpf	[ʃtʁʊmpf]
17. CCVC	Spiel	[ʃpil]	36. CCCVCCCC	pflanzt	[pflantst]
18. CCVVC	steil	[ʃtaɪl]	37. CCCVVCCC	streichst	[ʃtʁaɪxst]
19. CVCC	Gast	[gast]	38. CCVCCCCC	schrumpfst	[ʃʁʊmpfst]

Die Kombinationen der Konsonanten in diesen Silben weisen einige Beschränkungen auf. So können im Auslaut keine stimmhaften Obstruenten, /h/ oder /j/ auftreten; in Konsonantenhäufungen werden die folgenden Anforderungen berücksichtigt. Zweigliedrige Anlautverbindungen: Obstruent + Obstruent/Sonorant (bevorzugt Liquid); dreigliedrige Anlautverbindungen: Zischlaut + stimmloser Verschlusslaut + Liquid; mehrgliedrige Auslautverbindungen: Sonoritätsabfall wird in der Regel befolgt (an zweiter Stelle [s] oder [t] bevorzugt).¹⁰⁹

In gesprochener Sprache kommt es in unbetonten Silben zu Reduktionen, durch die vokalische Silbenkerne getilgt werden und konsonantische Silben entstehen:¹¹⁰ „-gen“ [gən] > [gŋ], „-gene“ [gnə] > [gŋnə]. Das Phänomen ist keineswegs auf die Umgangssprache beschränkt. Vokalausfall, viele komplexe Silbentypen und wort- sowie morphembezogene Regeln belegen, dass die Kategorie des phonologischen Wortes Vorrang vor der der Silbe hat. Die komplexen Konsonantenhäufungen lassen kein hohes angestrebtes Sprechtempo

¹⁰⁸ Zum Silbentyp 21: Die postvokalische konsonantische Aussprache von /ʁ/ mag als hyperkorrekt gelten, ist aber in den Nachrichtensendungen teilweise zu hören.

¹⁰⁹ Die knappe Zusammenfassung basiert auf den detaillierten Ausführungen in Hall (2000: 230 ff.).

¹¹⁰ Pompino-Marschall (2009: 275), nach Kohler (1990); auf das diakritische Zeichen zur Kennzeichnung silbischer Konsonanten wird hier aus optischen Gründen verzichtet.

vermuten. Der hier erhobene Wert für die mittlere Anzahl von Lauten/Silbe beträgt 2,99; der Anteil der CV-Silben an allen Silbentypen 24,7%.

Englisch:

Nach Cruttenden zählt das Englische einschließlich Affrikaten 24 Konsonanten- und mindestens elf Vokalphoneme:

Tabelle 4.13: Konsonantensystem des Englischen in Anlehnung an Cruttenden (2001: 149), hier ohne Affrikaten

	Bilabial	Labiodental	Interdental	Alveolar	Postalveolar	Palatal	Velar	Glottal
Plosiv	p b			t d			k g	
Nasal	m			n			ŋ	
Frikativ		f v	θ ð	s z	ʃ ʒ			h
Lateral				l				
Approximant						j	w	
Retroflex					ɻ			

Tabelle 4.14: Vokalsystem des Englischen in Anlehnung an Cruttenden (2001: 91)¹¹¹

i ɪ		u ʊ
e	ɜ	o
	(ə) ʌ	ɔ
æ		ɑ

Auch im Englischen ist die Aspirierung stimmloser oder ungespannter Verschlusslaute in fast allen Positionen außer nach /s/ als häufig zu bezeichnen (vgl. Spencer, 1996: 206 ff.). Neben den dargestellten Monophthongen verwendet das britische Englisch fünf „echte Diphthonge“ ([eɪ aɪ ɔɪ əʊ aʊ]) sowie drei weitere aus der Vokalisierung von /r/ entstehende zentrierende Diphthonge ([ɪə eə ʊə]). Gespannte und ungespannte Vokale weisen wie im Deutschen unterschiedliche Längen auf. Insgesamt beläuft sich die Zahl der Vokalphoneme damit auf maximal 20.

Von den Konsonantenphonemen werden 72,7% vorne gebildet, die 89,42% der artikulierten Konsonanten ausmachen:

Tabelle 4.15: Häufigkeit der englischen Konsonanten im gesprochenen Text

Artikulationsbereich	Laute	%	kumulativ
Labial	p b m f v	18,90	18,90
Koronal	θ ð t d n l s z ʃ ʒ ɹ	70,52	89,42
Dorsal	k g h ŋ	10,58	100,00

¹¹¹ Eine Diskrepanz zu IPA-Darstellungen stellt der Vokal [ʌ] dar, der dort als Hinterzungenvokal bezeichnet wird.

Im vokalischen Bereich können 33,33% der Phoneme als kürzer bezeichnet werden ([ɪ ʊ ɒ ə]).

Diese Phoneme können in relativ komplexen Silben miteinander kombiniert werden. Folgende Silbentypen treten im Englischen auf:¹¹²

Tabelle 4.16: Silbentypen im Englischen

Silbenstruktur	Bsp.	Aussprache	Silbenstruktur	Bsp.	Aussprache
1. V	a	[ə]	18. CCVVC	style	[staɪl]
2. VV	oh	[əʊ]	19. CVCC	hand	[hænd]
3. VC	an	[ən]	20. CVVCC	hound	[haʊnd]
4. VVC	owes	[əʊz]	21. CCCVC	splash	[splæʃ]
5. CV	to	[tə]	22. CCCVVC	strobe	[strəʊb]
6. CVV	go	[gəʊ]	23. CCVCC	brand	[brænd]
7. VCC	and	[ænd]	24. CCVVCC	clowns	[kləʊnz]
8. VVCC	owls	[aʊlz]	25. CVCCC	hands	[hændz]
9. CCV	flaw	[flɔ]	26. CCVVCCC	hounds	[haʊndz]
10. CCVV	blow	[bləʊ]	27. C ^c VCCCC	texts	[t ^h eksts]
11. CVC	bun	[bʌn]	28. CCVCCC	glimpse	[glɪmps]
12. CVVC	fowl	[faʊl]	29. CCVVCCC	crouched	[kraʊtʃt]
13. VCCC	ants	[ænts]	30. CCCVCC	springs	[sprɪŋz]
14. VVCCC	ousts	[aʊsts]	31. CCCVVCC	sprouts	[spraʊts]
15. CCCV	straw	[strɔ]	32. CCVCCCC	prompts	[prɒmpts]
16. CCCVV	spray	[spreɪ]	33. CCCVCCC	scripts	[skɪpts]
17. CCVC	grin	[grɪn]	34. CCCVCCCC	strengths	[streŋkθs]

Die häufig zitierten Triphthonge wie in „fire“ werden hier nicht berücksichtigt, da sie meist zu Diphthongen oder Monophthongen reduziert ([fa^ɪ]) oder, seltener, getrennt gesprochen werden (/faɪ.jə/). Auch im Englischen werden, jedoch zumindest in den Aufnahmen dieser Arbeit nicht so oft wie im Deutschen, Konsonantenhäufungen gebildet, in denen sich ein konsonantischer Silbenkern ausmachen lässt (/pi.pəl/ für „people“ ist häufiger als /pi.pl/). Für das Englische gilt in Bezug auf die angestrebte AG das für das Deutsche Gesagte. Die mittlere Anzahl der Laute/Silbe und der CV-Silbenanteil belaufen sich auf 2,81 und 21,8%.

¹¹² Der letzte Silbentyp stellt eine *mögliche* Aussprache dar und wird von Garn-Nunn/Lynn (2004: 100) angeführt. Weitere Silbentypen wären in einigen Varietäten durch eine konsonantische Artikulation von /r/ in der Coda denkbar, z.B. „arched“ /artʃt/. Im Britischen Englisch werden gespannte Vokale auch in nicht-akzentuierter Position häufig diphthongiert, z.B. „BBC“ etwa: [bjbj'sij].

Europäisches Portugiesisch:

Nach Morais Barbosa (1994b) besteht das Lautinventar des EP aus 21 Konsonanten (inkl. Gleitlaute) und acht Vokalen, die keinen phonologischen Längenkontrast aufweisen. Unter den Konsonanten befindet sich ein Phonem, das aus mehreren Artikulationsbewegungen besteht (/r/).

Tabelle 4.17: Konsonantensystem des EP nach Morais Barbosa (1994b: 139)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Palatal	Velar
Plosiv	p b		t d				k g
Nasal	m			n		ɲ	
Frikativ		f v		s z	ʃ ʒ		
Vibrant				r			
Getippt				ɾ			
Approximant						(j)	(w)
Lateral				l		ʎ	

Das EP weist eine Einschränkung der Vokalphoneme in unbetonten Silben auf. Die acht Vokalphoneme, die in betonten Silben vorkommen, lauten:

Tabelle 4.18: Vokalsystem des EP nach Morais Barbosa (1994a: 134)

i		u
e		o
ɛ	ɐ	ɔ
	a	

In finalen offenen oder durch Sibilanten geschlossenen Silben treten die Phoneme / a e u / auf, von denen /e/ und /u/ zur Nullrealisierung neigen; in durch Liquide oder Nasale geschlossenen finalen Silben treten / i e a o u / auf. In nicht-finalen Silben hinter der Wortbetonung kommen vor bedeckten Silben / i e a u /, vor nackten Silben / i u / vor. In vortonigen Silben schließlich können / i e a o u / stehen. Unbetonte Vokale werden qualitativ und quantitativ teilweise stark reduziert (s.u.). Zu den Monophthongen kommen nach Cunha/Cintra (2001: 47 f.) die Diphthonge [aj ej au eu ɨy oy ou ɔj ɔu uj ẽj ẽu õj õu] sowie die Triphthonge [waj wɛj wiw wẽw wɔj wõj]; Hiat und Diphthong drücken Bedeutungsunterschiede aus („pais“ [pajʃ] vs. „país“ [pa'is]). Fünf weitere Nasalvokale ([ĩ ã õ ẽ õ̃]) können als Allophone vor einem nasalen Archiphonem gesehen werden.¹¹³

¹¹³ Morais Barbosa (1994b: 136) fasst Argumente zusammen, darunter die plosive Artikulation von / b d g / nach /N/, die intervokalisches als Reibelauten [β ð ɣ] artikuliert werden; s. auch Mateus/Andrade (2000: 20 ff.).

Auch Morais Barbosa schließt für das EP öffnende Diphthonge aus und beschreibt „água“ als /a.gu.a/.

57,14% der Konsonantenphoneme werden im vorderen Artikulationsbereich gebildet, unter den Vokalen zählen 50% ([i u e o]) zu den kürzeren. Die vorne gebildeten Phoneme entsprechen 88,6% der artikulierten Konsonanten:

Tabelle 4.19: Häufigkeit der Konsonanten des EP im gesprochenen Text

Artikulationsbereich	Laute	%	kumulativ
Labial	p b m f v	19,20	19,20
Koronal	t d n l r s ʃ ʒ ɲ ʎ	69,40	88,60
Dorsal	k ɣ ʀ	11,40	100,00

Mateus/d'Andrade (2000: 40 ff.) weisen darauf hin, dass [i]¹¹⁴ in allen Positionen und [u] wortfinal in der Umgangssprache unrealisiert bleiben können, was in den Aufnahmen der vorliegenden Arbeit auch für Nachrichtensprache bestätigt werden kann, siehe Abbildung 4.2.

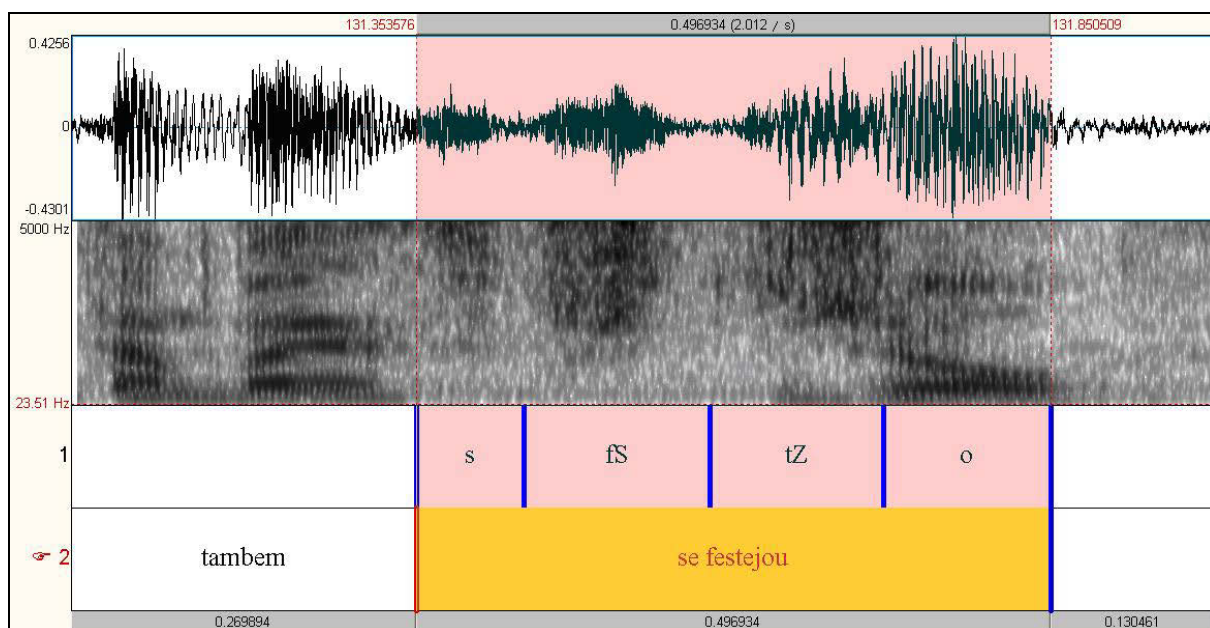


Abbildung 4.2: Vortönige Nullrealisierung von /e/ in EP "se festejou"

Die folgende Liste stellt eine um die in Mateus/d'Andrade (2000: 43) und eigene Beispiele ergänzte Liste nach Morais Barbosa (1994a: 137) dar:

¹¹⁴ Mateus/d'Andrade (2000) transkribieren [i] an Stelle von [ə].

Tabelle 4.20: Silbentypen im EP mit den entsprechenden Silben in mehrsilbigen Wörtern fett gedruckt¹¹⁵

Silbenstruktur	Bsp.	Aussprache	Silbenstruktur	Bsp.	Aussprache
1. V	é	[ɛ]	11. CVVCC	quais	[kwajʃ]
2. VC	és	[ɛʃ]	12. CCVVC	tren	[trẽ̃ɲ]
3. VVC	eis	[eɪʃ]	13. CCVVCC	trens	[trẽ̃ɲ]
4. CV	má	[ma]	14. CCCV	esdrú xula	[ʒdru]
5. CVV	tem	[tẽ̃]	15. CCCVC	telefone	[tlfɔ̃]
6. CVC	mal	[maɫ]	16. CCCCVC	festejou	[fɛʃtʒo]
7. CVVC	pais	[paɪʃ]	17. CCCCVC	despegar	[dɛʃpgar]
8. CCV	cru	[kru]	18. CCCCVC	despregar	[dɛʃprgar]
9. CCVV	pleito	[plẽ̃]	19. CCCCCVC	desprevenir	[dɛʃprvnir]
10. CVVC	qual	[kwaɫ]			

Phonologisch leuchten die Argumente von Mateus/d'Andrade (2000: 44f.) ein, dass in Konsonantenclustern, die aus Vokalelision entstehen, leere Silbenkerne vorliegen und diese Beispiele als Mehrsilber analysiert werden müssen. Es ist nicht von der Hand zu weisen, dass phonetisch höchst komplexe Konsonantenhäufungen entstehen können. Da die genannten Konsonantengruppen durch Vokalausfall unbetonter Vokale entstehen und im EP die Betonung maximal auf die drittletzte Silbe fallen kann, sind umfangreiche Cluster eher im Wortanlaut möglich.¹¹⁶ Silbenpräferenzgesetze sind im EP kaum von Bedeutung, da es sich um eine relativ strikte Wortsprache handelt: Silbenauslautende Konsonanten werden nicht wie im BP vokalisiert (EP [maɫ] vs. BP [mau] für <mal>, de. „schlecht“). Beispiele wie [prtɛzɫ] (für <portugueses>, de. „portugiesische“) scheinen ein Bestreben zu belegen, Wortränder näher aneinander zu bringen, zwischen denen nur ein vokalischer Kern auftritt. Gleichzeitig können Konsonanten, die einen hohen Grad an Sonorität aufweisen, als Silbenkerne bezeichnet werden. [fɛ] und [pɾt] können dann als konsonantische Silben bezeichnet werden, ähnlich wie in en. „little“ [litl] und de. „laufen“ [laufm], wobei es sich bei diesen konsonantischen Silbenkernen um stimmhafte Sonoranten handelt. Ob [fɛ] und [r] tatsächlich phonetische Eigenschaften besitzen, die ihnen eine Funktion als Silbenkern zukommen lassen können, ist zweifelhaft. Im Falle von [fɛ] scheint die notwendige phonetische Substanz aufgrund mangelnder Stimmhaftigkeit nicht erreichbar, bei [r] handelt es sich zwar um einen stimmhaften Liquid, der in einigen Sprachen als silbenkernfähig bezeichnet wird,¹¹⁷ ist aber als geschlagener Laut von derart

¹¹⁵ Im Silbenkopf liegen für die Kombination von Konsonantenphonemen Beschränkungen vor: Der zweite Konsonant muss ein Liquid oder Nasal sein, im Silbenauslaut kommt nur die Kombination / n s / vor.

¹¹⁶ Im Gegensatz zum Chinesischen.

¹¹⁷ Rehder (2006: 333) erwähnt unter den slawischen Sprachen, in denen /r/ als konsonantischer Silbenkern bezeichnet wird, nur für das Makedonische eine mögliche Aussprache mit vokalischem Übergang: „Neben

kurzer inhärenter Dauer, dass eine Reduktion und schließlich der totale Ausfall begünstigt scheinen (vgl. es. „para“ > „pa“, de. „für, um zu“). In [fj] liegt keine Öffnungs- und Schließbewegung vor, die Sonorität weist keinen nennenswerten Unterschied auf. Ungeachtet der phonologischen Bewertung derartiger Phänomene soll hier festgehalten werden, dass komplexe Konsonantenhäufungen im EP auch im formellen Redestil vorkommen.

In den portugiesischen Nachrichtensendungen wurden durchschnittlich 2,38 Laute/Silbe und ein Anteil von 31,49% CV-Silben gezählt. Dass dieser Anteil höher liegt als im BP kann darauf zurückgeführt werden, dass im BP Diphthongierungen Abweichungen von der CV-Struktur generieren: „Pomba“, de. „Tauben“, besteht phonologisch aus zwei CV-Silben /põ.ba/, wird aber als [põ̃ba] gesprochen. Andererseits haben die Vokalreduktionen im EP oft einen zwar qualitativ wie quantitativ stark reduzierten, aber dennoch vorhandenen vokalischen Übergang zur Folge („Cavaco“ /ka'vako/ [kʰvak]). Die erwähnten Tendenzen der beiden Varietäten (BP „fnac“ > [fi'naki], EP „festejou“ > [fʃtʃo], de. „feierte“) sind offensichtlich im untersuchten Material nicht so stark ausgeprägt, dass sie sich im CV-Silben-Anteil bemerkbar machen würden.

Spanisch:

Quilis (2005) zählt die folgenden 19 Konsonantenphoneme auf:¹¹⁸

Tabelle 4.21: Konsonantensystem des Spanischen in Anlehnung an Quilis (2005: 86 f.)

	Bilabial	Labiodental	Interdental	Dentoalveolar	Alveolar	Palatal	Velar
Plosiv	p b			t d			k g
Nasal		m			n	ɲ	
Frikativ		f	θ		s		x
Affrikativ						tʃ	
Vibrant					r		
Getippt					ɾ		
Approximant						j	
Lateral					l	ʎ	

der silbischen Aussprache ohne Begleitvokal [ɾ] findet sich fakultativ [əɾ].“ Für Tschechisch, Slowenisch, Slowakisch, Kroatisch und Serbisch wird in den Beiträgen zum Sammelband nicht von einem Sprossvokal [ə] ausgegangen. Vgl. Comrie/Corbett (1993), Press (1986).

¹¹⁸ Die Verwendung des Symbols ‘tʃ’ für die palatale stimmlose Affrikate ist zwar IPA-konform, jedoch deshalb nicht sinnvoll, weil [t] im Spanischen nicht palatal, sondern dental artikuliert wird und kein Laut /ʃ/ existiert.

Im Vokalsystem werden fünf phonologisch gleich lange Phoneme unterschieden:¹¹⁹

Tabelle 4.22: Vokalsystem des Spanischen nach Quilis (2005: 37 f.)

i		u
e		o
	a	

Davon können 80% ([i u e o]) phonetisch als potentiell kürzer bezeichnet werden. Das Spanische kennt sowohl öffnende und schließende Diphthonge ([ja je jo wa we wo aɪ eɪ oɪ aʊ eʊ oʊ]) als auch Triphthonge ([jaɪ jeɪ waɪ weɪ], Wortgrenzen überschreitend auch weitere).

Quilis (1999: 79) merkt an, dass die Ausgangsposition für die Artikulation im Spanischen vorne im Mundraum liegt („La Lengua española general tiene una base de articulación anterior, sonora, deslabializada, creciente, tensa, y de acción glótica normal“).¹²⁰ Insgesamt werden nach seinen Angaben 68,58% der Phoneme des Spanischen im vorderen Bereich des Mundraumes artikuliert, darunter 46,53% von den Konsonanten (8,23% Labiale, 11,08% Dentale, 25,81% Alveolare und 1,41% Präpalatale) und 22,05% von den Vokalen. Die Zahlen scheinen die Häufigkeit der Laute zu berücksichtigen („... el 60,58% de todas nuestras articulaciones son anteriores“, ebd.), wobei die angegebene Häufigkeit von /i/ und /e/ zusammen mit nur 22,05% recht gering scheint. In den Auswertungen der Nachrichtenaufnahmen ergibt sich folgende Verteilung konsonantischer Laute:

Tabelle 4.23: Häufigkeit spanischer Konsonanten im gesprochenen Text¹²¹

Artikulationsbereich	Laute	%	kumulativ
Labial	p b m f v	16,28	16,28
Koronal	t d n l r s ʃ ʒ	73,28	89,56
Dorsal	k g x	10,44	100,00

Die 52,38% der Konsonantenphoneme, die vorne artikuliert werden, werden im gesprochenen Text häufig benutzt und stellen 89,56% der produzierten Konsonanten dar. Von 19 Konsonantenphonemen bestehen zwei aus mehreren Artikulationsbewegungen: die

¹¹⁹ Hier werden die in der Hispanistik verwendeten Symbole ‘e’ und ‘o’ für die halboffenen Vokale übernommen, was im Grunde bezüglich des Öffnungsgrades irreführend ist.

¹²⁰ Man bedenke, dass wichtige Unterschiede bezüglich der Ruheposition der Artikulationsorgane auch zwischen Varietäten bestehen, vgl. Wängler (1983: 167).

¹²¹ [ʃ] wird hier als zweiter Bestandteil der Affrikate [tʃ] gezählt (s.u.); [v] und [ʒ] in Fremdwörtern.

Affrikate /tʃ/ und der Vibrant /r/, der aus mindestens zwei Schlägen (kurzen Verschlüssen) besteht.¹²²

Die möglichen Silbentypen des Spanischen sind bei Quilis¹²³ auf neun Strukturen begrenzt, von denen die komplexeste fünf Laute umfasst und in der gesprochenen Sprache 0,01% aller Silben darstellt. Mit 55,81% der gesprochenen Silben ist die CV-Silbe der häufigste Silbentyp und tritt mehr als doppelt so oft auf wie die einfache, geschlossene Silbe (CVC), die mit 21,61% der Silben den zweiten Rang einnimmt.¹²⁴ In Anlehnung an Colina (2009: 11) und Canellada/Kuhlmann Madsen (1987: 44 f.) gilt diese umfassende Liste:

Tabelle 4.24: Silbentypen im Spanischen mit den entsprechenden Silben in mehrsilbigen Wörtern fett gedruckt

Silbenstruktur	Bsp.	Aussprache	Silbenstruktur	Bsp.	Aussprache
1. V	o	[o]	10. VV	hay	[aj̥]
2. CV	tu	[tu]	11. CVV	soy	[soj̥]
3. CVC	por	[por]	12. VVC	aunque	[aũj̥]
4. VC	en	[en]	13. CVVC	vais	[baj̥s]
5. CCV	tragar	[tra]	14. CVVV	buey	[bwej̥] ¹²⁵
6. CCVC	tres	[tres]	15. CCVV	pleitear	[plej̥]
7. VCC	instaurar	[ins]	16. CCVVC	claustró	[klaũs]
8. CVCC	vals	[bals]	17. CVVVC	cambiáis	[bjaĩs]
9. CCVCC	transporte	[trans]	18. CCVVVC	ampliáis	[pljaĩs]

Auch im Spanischen kann es zu Reduktionen unbetonter Vokale kommen, wobei derartige Beobachtungen in der Regel in hispanoamerikanischen Varietäten gemacht werden.¹²⁶

Konsonantenhäufungen unterliegen starken Beschränkungen, so lassen sich im Silbenauslaut nur Gruppen bilden, deren zweiter Konsonant /s/ ist. Auslautende Konsonanten im Wortinneren sind für Neutralisationen anfällig (/p b t d ɡ k / > / B D G /), am Wortende kommen im indigenen Wortschatz keine Konsonantenhäufungen vor. Im Wortanlaut können nur ein Verschlusslaut (phonologisch betrachtet) bzw. /f/ mit einem Liquid kombiniert werden (Berschin et al. 2005: 144). Wortränder werden durch Resilbifizierung verwischt und Silbenränder gestärkt.¹²⁷ Es wird deutlich, dass das Spanische von den hier

¹²² Quilis (2005: 65); nach Canellada/Kuhlmann Madsen (1987: 42) mindestens drei Schläge, bei Schwegler et al. (2010: 278) zwei oder drei Schläge.

¹²³ Quilis (1993: 370), gestützt auf Daten aus Guerra (1983).

¹²⁴ Werte aus Berschin et al. (2005: 145), die sich auf Lloyd/Schnitzler (1967) und Delattre (1965) stützen.

¹²⁵ [w] steht hier, wie in Quilis (2005), für einen labiovelaren Halbkonsonanten (vgl. Schwegler et al., 2010, die [w] für Konsonanten, [ɥ] für Halbkonsonanten und Halbvokale verwenden).

¹²⁶ S. Lipski (1990) zu ecuadorianischem Spanisch; Lope Blanch, (1983: 57 ff.), zu mexikanischem Spanisch; Delforge (2008a,b) zu peruanischem Spanisch.

¹²⁷ „Las aves“ (de. „die Vögel“), „la sabes“ (de. „du kennst sie / weißt es“): [la.ˈsa.βes] (Canellada/Kuhlmann

behandelten Sprachen die strikteste Tendenz zur CV-Struktur aufweist und am häufigsten vordere Konsonanten verwendet. Die Silbenkomplexität liegt bei 2,34 Lauten/Silbe und die Häufigkeit der CV-Silben bei 33,09%.

Ein Vergleich der Phoneminventare und Konsonantenhäufigkeiten zeigt, dass die kleinsten Vokalsysteme im Spanischen und Chinesischen vorliegen, das Deutsche weist mit 15 Vokalphonemen das umfangreichste System auf. Die Häufigkeit der Konsonanten fällt vor allem im Spanischen mit 89,56% zugunsten potentiell kürzerer Laute aus, wobei Englisch (89,42%) und EP (88,60%) etwa im gleichen Bereich liegen. Eine eindeutige Präferenz der einzelnen Sprachen für kurze Laute, die eine hohe Lautgeschwindigkeit begünstigen, kann damit nur für das Spanische angenommen werden, das als Sprache mit dem kleinsten Vokalinventar auch keine Langvokale aufweist. Streng als quantitative Kontraste zu wertende Oppositionen im vokalischen oder konsonantischen Bereich werden von keiner der hier behandelten Sprachen phonologisch genutzt, wobei in den germanischen Sprachen gespannte Vokale deutlich länger sind als ungespannte. Viele komplexe Silben, die die Silbendauer erhöhen und die Silbengeschwindigkeit mindern, sind im Deutschen und Englischen zu erwarten.

4.3 Material

Um Vergleichbarkeit zwischen den hier untersuchten Einzelsprachen zu gewährleisten wird das Sprachmaterial auf eine Sprechsituation beschränkt, die in den Sprachen (bzw. Kulturen) ähnliche Ausprägungen annimmt: vorbereitetes Vorlesen. Material, das durch Spontaneität gekennzeichnet ist, eignet sich deshalb nicht, weil der Charakter einzelner Sprecher zu stark zur Geltung kommt und eine unrealistisch große Stichprobe vorliegen müsste um mit dem Idiolekt zusammenhängende Effekte ausgleichen zu können. Des Weiteren sollen diskursrelevante und pragmatische Einflussfaktoren sowie Faktoren wie die emotionale Verfassung und die Phrasenlänge weitgehend ausgeschlossen werden. Letztere wird in vorbereiteten informativen Texten, die leicht verständlich und zugleich formell wirken sollen, als relativ normiert angenommen. Als Sprachmaterial werden Nachrichtensendungen untersucht, die aus Podcasts im Internet aufgenommen wurden. Die tendenzielle Vermeidung großer lokaler Veränderungen des Tempos bei vorbereitetem Vorlesen in einer formellen Sprechsituation ermöglicht außerdem einen verlässlichen

Madsen, 1987: 43); „los amigos” (de. „die Freunde“): [lo.sa.'mi.yos] (Colina, 2009: 45); es handelt sich beim Spanischen um eine typische Cursus-Sprache im Sinne Pulgrams (1970).

Durchschnittswert des globalen Sprechtempos. Die Wahl des Sprachmaterials ermöglicht, dass die meisten genannten Faktoren so kontrolliert gehalten werden, dass die Auswertung spezieller Faktoren wie Sprache und Geschlecht in den Vordergrund gestellt werden kann. Die Nachrichtensendungen wurden so gewählt, dass hauptsächlich die nationale Standardnachrichtensendung als Ausgangsmaterial dient.¹²⁸ Hierdurch wird die „angesehene Norm“ einer Sprache als Gegenstand der Betrachtung festgelegt und regionale Varietäten werden ausgeschlossen. Diese werden keineswegs als nicht gleichwertig angesehen, aber aufgrund ihrer begrenzten Verbreitung und weniger stark ausgeprägten Normierung als nicht angemessen für diese Art von Untersuchung verstanden. Da neben der phonetischen Untersuchung auch kulturelle Aspekte im Vordergrund der Arbeit stehen, sind diese Sendungen eher als regionale und/oder durch Spontansprache ausgezeichnete Aufnahmen geeignet um kulturelle Normen und sogenannte *cultural scripts* (s. 7.4.3) auszumachen. Der Einwand, es handle sich bei dieser Art von Sprachmaterial um einen unrealistischen Untersuchungsgegenstand, der eine unnatürliche Verwendung der Sprache darstellt, kann zurückgewiesen werden, da aufgrund der notwendigen Vergleichbarkeit und Repräsentativität kein anderes Material in Frage kommt (vgl. hierzu das Material von Meinhold, 1967, aus dem hervorgeht, dass die Streubreite der untersuchten Variablen bei Nachrichtentexten deutlich geringer ausfällt als bei den anderen behandelten Textsorten, was den Sprachvergleich erleichtert). Vorbereitetes Vorlesen stellt keine unnatürliche, sondern eine vorbereitete Art der Kommunikation von Informationen in einer natürlichen Sprechsituation dar. Vgl. hierzu Benton (2010), der argumentiert, dass Nachrichtensprache zumindest gegenüber der „Laborsprache“ den großen Vorteil der Natürlichkeit mit sich bringt und gleichzeitig die für eine Sprachgemeinschaft als vorbildlich geltende Norm darstellt. Hier wird davon ausgegangen, dass die Nachrichtensprache als ein ästhetisches Ideal zu verstehen ist, das auf die alltägliche Kommunikation jedoch keine präskriptive Wirkung ausübt. Allerdings darf die Rolle der Medien bei der Vermittlung von sprachlichen Normen nicht unberücksichtigt bleiben. Zwar ist vermutlich der Einfluss von Nachrichtensprache bei der Sozialisierung von Kindern gering, dennoch sollte erwähnt werden, dass z.B. Manusov/Jaworski (2006) davon ausgehen, dass Medien zumindest in Bezug auf die Aneignung nonverbaler Verhaltensweisen einen nachhaltigen Einfluss haben. Die Ergebnisse der vorliegenden empirischen Studie müssen selbstverständlich auf das untersuchte Register und die entsprechende Redesituation eingeschränkt werden. Trouvain (2004) hält zusammenfassend fest, dass in Nachrichtensprache im Allgemeinen eine relativ

¹²⁸ Die einzelnen Sendungen werden in den folgenden Unterkapiteln genannt, in denen die Ergebnisse zusammengefasst werden.

hohe SG bevorzugt wird (gestützt auf Ergebnisse von Fackrell et al., 2000; Meinhold, 1967 und Fónagy & Magdics, 1960).¹²⁹ Damit muss berücksichtigt werden, dass die in dieser Arbeit gemessenen Werte möglicherweise im oberen Bereich des Streubereichs von für Einzelsprachen typischen Sprechtempi liegen.

4.4 Methodik

4.4.1 Aufnahme und Transkription

Für alle Sprachen wurden Aufnahmen von sechs bis acht Sprechern pro Geschlecht untersucht. Die Aufnahmen wurden mit dem Programm *Audacity* direkt am Computer aufgenommen und im *Wave*-Format gespeichert. Einige Dateien mussten zur Reduzierung der Hintergrundgeräusche bzw. der Hintergrundmusik mit den in *Audacity* zur Verfügung gestellten Funktionen nachbearbeitet werden. In den meisten Fällen wurde der Gesamtnachrichtenüberblick zu Beginn einer Sendung aufgenommen, um vergleichbare Voraussetzungen zu schaffen. Das Zusammenschneiden aus verschiedenen Beiträgen über die gesamte Sendung hinweg wurde vermieden, um mögliche Einflüsse der Länge der Äußerungen weitgehend zu vermeiden. Die Dauer dieser einzelnen Moderationen weist sehr große Unterschiede auf. Für einige Sprachen (Brasilianisch, Deutsch, Englisch, Spanisch, teilweise Chinesisch) war es möglich, Sendungen oder Teile von Sendungen zu verwenden, die aus einem an einem Stück gesprochenen Nachrichtenüberblick bestehen.

Die Dateien wurden daraufhin mehrmals abgehört und zunächst orthografisch, dann phonetisch breit transkribiert. Die breiten Transkriptionen wurden vom Autor der Arbeit sonagramm- und oszillogrammgestützt zu detaillierteren engen Umschriften erweitert, wozu das Programm *Praat* verwendet wurde. Anschließend wurden Laut- und Silbenzahlen erhoben. Es wurde großer Wert darauf gelegt, mit der Transkription ein möglichst realistisches Abbild der tatsächlich realisierten Einheiten zu liefern. Dem Glottisschlag [ʔ] und dem Hauchlaut [h] wurden Lautstatus bei der Auszählung zugerechnet. Vorvokalische Gleitlaute (Semikonsonanten) wurden zu den Konsonanten, nachvokalische (Semivokale) zu den Vokalen gezählt, um den konsonantischen Anteil an allen artikulierten Lauten zu berechnen. Zur Abgrenzung von Lauten wurden u.a. Hinweise in Pompino-Marschall (2009) und Machelett (1996) herangezogen. Die Transkriptionen wurden in einem Textverarbeitungsprogramm erstellt um die anschließende Erhebung der

¹²⁹ Hegedüs (1957: 15) meint jedoch, Nachrichtensprache erscheine dem Hörer im Ungarischen „von mittlerem Tempo“.

Laut- und Silbenzahlen zu vereinfachen. Die Umschrift erfolgte in Anlehnung an die SAM-PA-Transkriptionskonventionen (*Speech Assessment Methods Phonetic Alphabet*). Die nicht ganz unvermeidbaren (qualitativen) Ungenauigkeiten dieser Transkription sind für die Untersuchung unerheblich, da die Anzahl der Laute, nicht die exakte qualitative Beschreibung im Fokus steht. Darüber hinaus enthält natürlich jede Transkription als idealisierte Abfolge diskreter Segmente einen hohen Abstraktionswert. Zweifelsfälle wurden mit linguistisch ausgebildeten Muttersprachlern besprochen.

4.4.2 Definitionen der ausgezählten sprachlichen Elemente

4.4.2.1 Definition von Lauten als zählbare sprachliche Elemente

Für die Einheit Laute/s ist bei der Auszählung problematisch, dass einzelne Laute oder Segmente in gesprochener Sprache nicht isolierbar sind, da sie immer einen bestimmten Grad an Koartikulation mit Nachbarlauten aufweisen. Unter dem Begriff „Laut“ wird hier ein Lautereignis verstanden, das als akustisches Ergebnis des Erreichens oder Anstrebens einer artikulatorischen Zielkonfiguration (*Target*) vorliegt. Das *Target* weist in gesprochener Sprache unterschiedliche, nicht invariante Realisierungen auf und wird oft nicht erreicht. Einzelne Lautereignisse können auf verschiedene artikulatorische Konstellationen zurückgeführt werden (Atal et al., 1978 zu unterschiedlichen Beschaffenheiten des Vokaltraktes bei der Artikulation von Vokalen). Das *Target* eines stimmhaften dentalen Plosivs /d/ z.B. ist ein Verschluss zwischen Zungenspitze und einem Bereich an den Schneidezähnen (je nach lautlicher Umgebung kann der Punkt in der Horizontalen verschoben werden). Wenn bei der Realisierung von /d/ aufgrund von Beschränkungen im Bereich der zeitlichen Ausdehnung der Artikulationsbewegung oder des artikulatorischen Aufwandes, der erbracht wird, an Stelle eines Verschlusslautes ein Reibelaut oder ein geschlagener Laut vorliegt, ist das als eine Artikulationsbewegung zu werten, die zwar nicht zum *Target* des entsprechenden Phonems führt, aber als das intendierte Phonem wahrgenommen werden kann (sofern keine Interferenz mit anderen Phonemen auftritt und die Realisierung z.B. als Versprecher gewertet oder die lexikalische Einheit missverstanden wird). Das Lautereignis, in diesem Fall z.B. [ð], wird dann als Laut gezählt. Es handelt sich hierbei um eine Art von *undershoot*-Realisierung, die z.B. aus Gründen der Sprachökonomie als solche produziert wird (vgl. Lindblom, 1983). Außerdem kann es sein, dass zugrunde liegenden Phonemen bei der Artikulation keine Realisierung entspricht, also kein Lautereignis im Sprachsignal nachweisbar ist, in welchem Fall kein

Laut gezählt wird.¹³⁰ Laute, die aus komplexen Bewegungsabläufen bestehen, wurden als Summe ihrer Bestandteile gewertet. Affrikaten, z.B. es. [tʃ], wurden daher als zwei Laute gezählt. Auch behauchte Verschlusslaute wurden als zwei Laute gezählt, da zwei Bereiche im akustischen Sprachsignal und beim Höreindruck nachweisbar sind.¹³¹ Anders als bei /h/ ist für die ausströmende Luft bei behauchten Verschlusslauten ein Artikulationsort festzumachen, an dem die Luft aufgestaut wird und an dem ein Moment der Reibung zu erkennen ist.¹³² Diese Behauchung ist in den behandelten germanischen Sprachen meist vorhanden, in den romanischen Sprachen dagegen normalerweise nicht zu beobachten.¹³³ Auch Diphthonge wurden, auch wenn sie artikulatorisch nicht in zwei trennbare Ereignisse zerlegt werden können, als zwei Laute gezählt, da in der Zählung der Unterschied zu Monophthongen, die an derselben Stelle artikuliert werden könnten, berücksichtigt werden soll.¹³⁴ Somit liegt den Lautelementen eine auch auf artikulatorischen, hauptsächlich aber akustischen Aspekten beruhende Definition zugrunde. Bei der Zählweise von Lauten wurde außer Acht gelassen, dass einzelne Muskelgruppen der Artikulationsorgane im Zusammenwirken ergeben, was als Abfolge von Lautereignissen idealisiert werden kann (Liberman et al. 1967: 446 f.).

¹³⁰ Auf die Diskrepanz zwischen tatsächlicher Lautzahl und kanonischer (=phonemischer) Lautzahl wird in Kap. 5 näher eingegangen.

¹³¹ Dabei wird stark vereinfachend nur zwischen Vorhandensein und Nicht-Vorhandensein der Behauchung unterschieden, die jedoch unterschiedlich stark ausfallen kann. Im Fall von behauchten Reibelauten wird starke Behauchung nicht als separate Artikulationsbewegung gewertet, da sie im Gegensatz zur Behauchung von Verschlusslauten als beschreibendes Merkmal der produzierten Reibung selbst aufgefasst werden kann. Behauchte und unbehauchte Verschlusslaute können durchaus, bezugnehmend auf den verzögerten Stimmtoneinsatz und den Zeitpunkt der maximalen Glottisöffnung, durch die eine Schwingung der Stimmbänder vermieden wird, als längere oder kürzere Segmente gewertet werden (vgl. Dogil, 2003: 546 f.). Behauchung zählt in diesem Sinne als Merkmal *eines* Lautes. Um der geräuschhaften Transition im Sprachvergleich Rechnung zu tragen, wird dieses „Mehr“ im akustischen Bereich hier wie ein eigenes Segment behandelt, wobei bei der Präsentation der Ergebnisse der empirischen Studie auch darauf eingegangen wird, welche Verhältnisse sich bei der traditionellen Betrachtung behauchter Verschlusslaute als ein Laut ergeben. Damit wird eine zeitlich spätere maximale Glottisöffnung bei stimmlosen behauchten Verschlusslauten gegenüber unbehauchten quantitativ berücksichtigt. Ähnliches gilt für die Affrikate [tʃ] im Spanischen, die als ein Segment beschrieben werden kann, aufgrund der zeitlich unflexiblen Abfolge von Verschluss und folgender Reibung jedoch als zwei Lautereignisse gewertet werden soll.

¹³² S. hierzu Kohler (1995: 158).

¹³³ Im EP ist diese Behauchung in manchen Fällen zu hören.

¹³⁴ Dabei ist u.a. an Diphthongierungen (BP „mas“ > [maj]) und Monophthongierungen (es. „fuego“ > [ˈfojo]) und „Minimalpaare“ (de. „Hass“ / „Haus“) zu denken.

4.4.2.2 Definition von Silben als zählbare sprachliche Elemente

Nicht unproblematisch ist auch die Definition der Silbe. Geht man davon aus, dass ein vokalischer Silbenkern das Kriterium zur Einteilung von Äußerungen in Silben ist, verringert sich die Anzahl der unter anderen Gesichtspunkten wie dem des Sonoritätsverlaufs gezählten Silben in Sprachen wie dem Deutschen bei gleichzeitigem Anstieg der durchschnittlichen Anzahl von Lauten pro Silbe, wie folgendes Beispiel verdeutlicht: „Lebensmittel“ wird durch Vokalelision als Zweisilber [le:^bms.mɪtl] mit jeweils fünf bzw. vier Lauten pro Silbe realisiert. Vokalische Silbenkerne eignen sich für die phonologische Einteilung in Silben; für eine phonetische Abgrenzung von Silben wurden Öffnungs- und Schließbewegungen, Druckimpulse (Stetson, 1928) und der Sonoritätsverlauf (Sievers, 1901) herangezogen, aber auch die Domäne der Koartikulation (vgl. Kozhevnikov/Chistovich, 1966) - allerdings kann kein Kriterium allgemein für eine Einteilung in phonetische Silben anerkannt werden. Für das erwähnte Beispiel können nach dem Kriterium der Öffnungs- und Schließbewegungen mindestens zwei Öffnungsbewegungen (für [e:] und [ɪ]) ausgemacht werden.¹³⁵ Kohler (1995: 74) betont die geringere Intensität von [ɪ] im Sprachsignal im Vergleich zu manchen Sibilanten, was eine alternative Silbenzählung von drei Silben in [le:^bms.mɪtl] möglich erscheinen lässt. Nach dem Kriterium der Sonorität liegen vier Sonoritätsgipfel vor ([e: m ɪ l]), die jeweils von weniger sonoren Lauten umgeben sind ([le:^bms.mɪ.tl]). Kozhevnikov/Chistovich (1966: 128 f.) verstehen unter der phonetischen Silbe als artikulatorischer Einheit eine Abfolge von Sprachlauten, die auf einen Vokal endet. Alle davor artikulierten konsonantischen Laute zählen sie zu dieser artikulatorischen Silbe, folgende Konsonanten bilden einen Teil der folgenden Silbe. Diese von den Autoren als kleinste artikulatorische Einheit der Sprache bezeichnete Definition phonetischer Silben wird in einigen Untersuchungen übernommen.¹³⁶ Allerdings sind die Daten, auf die sich Kozhevnikov und Chistovich stützen, nicht ohne Zweifel zu betrachten: Drei Sprecher des Russischen produzieren bedeutungslose Logatome und „echte“ Wörter, wobei nur Konsonantengruppen von zwei Segmenten zwischen Vokalen auftreten. Bei einem der Sprecher sind die Ergebnisse nicht aussagekräftig; bei den anderen beiden kann ein hoher Grad an Koartikulation zwischen Konsonanten und folgenden Vokalen festgestellt werden. Dabei weist der Vokal eine assimilatorische Wirkung auf die beiden vorausgehenden Konsonanten aus, auch, wenn

¹³⁵ Bei [s] muss der Kieferabstand aber nicht geringer sein als bei [ɪ], vgl. Tillmann/Pfützinger (2003).

¹³⁶ In leicht abgeänderter Form u.a. in Dellwo (2010).

eine phonemische Silbengrenze zwischen ihnen liegt. Wie jedoch aus anderen Untersuchungen hervorgeht (z.B. Gafos et al., 2010; Koreman/Morland, 2007; Magen, 1997) ist der Grad der Koartikulation u.a. sprecherabhängig.¹³⁷ Die Daten von Kozhevnikov/Chistovich sollten damit nicht als Beweis für eine allgemeingültige Struktur „C_nV“ artikulatorischer Silben verstanden werden. Des Weiteren weist Öhmann (1966) anhand akustischer Daten nach, dass in VCV-Sequenzen mit medialem Verschlusslaut im Schwedischen (zwei Sprecher) und im amerikanischen Englisch (ein Sprecher), nicht aber im Russischen (ein Sprecher) der hintere Vokal bereits mit dem vorderen, also vor dem Konsonanten, koartikuliert wird und damit eine nicht-lineare artikulatorische Einheit vorliegt (vgl. Recasens, 1987, der für Spanisch und Katalanisch eine Abhängigkeit des Grades der V₁-V₂-Koartikulation in V₁CV₂-Sequenzen von den artikulatorischen Beschränkungen durch C und V₂ für progressive und durch V₁ für regressive Assimilation feststellt). Da innerhalb dieser Einheit eine Silbengrenze verläuft, kann die Silbe nicht uneingeschränkt als artikulatorische Einheit, die über Koartikulation definiert wird, angenommen werden (vgl. auch Harris, 1977: 76 f., die davon ausgeht, dass Silbengrenzen koartikulatorische Prozesse nicht blockieren; aber Gay, 1977, der in seinen artikulatorischen Daten Koartikulation in VCV-Sequenzen praktisch nur zwischen C und V₂ feststellt;¹³⁸ Magen, 1997, liefert Hinweise, dass im Englischen koartikulatorische Effekte auf der Fußebene über mehrere Silben hinweg auftreten können). In dem hier vorgeführten Beispiel scheint eine Silbeneinteilung in [le:^bmsmɪ.tl] weder artikulatorisch noch phonologisch plausibel.

Nach unterschiedlichen Zählweisen ergeben sich zwischen zwei und vier Silben für das deutsche Wort „Lebensmittel“. Da nur manche Sprachen Laute in artikulatorischen Einheiten zusammenzufassen scheinen, in denen kein vokalisches Element enthalten ist, alle Sprachen jedoch über „Sprechsilben“ mit vokalischem Silbenkern verfügen, liegt es nahe, diesen zum allgemeingültigen Kriterium zu erklären. Dieser Ansatz ist auf den ersten Blick nicht in der Lage, der Tatsache Rechnung zu tragen, dass Sprachen unterschiedliche Gruppierungen in vokalische und konsonantische Elemente vornehmen und manche Sprachen konsonantische Silben zulassen. Diese strukturellen Unterschiede können aber berücksichtigt werden, wenn die durchschnittliche Anzahl von Lauten pro Silbe verglichen

¹³⁷ Kent (1983) gibt außerdem einen Hinweis darauf, dass der Grad der Koartikulation, genau wie die AG, vom Alter des Sprechers abhängt.

¹³⁸ Gay (1978) spricht für ein auf Segmenten beruhendes Modell der Sprachproduktion und zeigt, dass das Ausmaß (die Anzahl der betroffenen Segmente) regressiver Assimilation von der Anzahl der Segmente in der vorausgehenden phonetischen Kategorie (C bzw. V) abhängt.

wird: Sprachen mit einer hohen Anzahl von Lauten pro Silbe (mit vokalischem Kern) lassen auf eine Tendenz zur Toleranz konsonantischer Silben schließen. Auf eine alternative Silbenzählung unter Berücksichtigung konsonantischer Silben wird bei der Präsentation der Ergebnisse auch eingegangen.

4.4.2.3 Definition von Morphemen¹³⁹ als zählbare sprachliche Elemente

Als Morphem wurden nach Bloomfield nicht weiter teilbare bedeutungstragende Lautfolgen gezählt.¹⁴⁰ Gleichzeitig wurden nur lautlich umgesetzte Morpheme gezählt. Nullmorpheme, wie sie u.a. von Lemos Monteiro (2002) in pt. „autor“: {autor}, {Ø} als Markierung des männlichen Geschlechtes in Kontrast zu „autora“: {autor}, {a} analysiert werden, blieben unberücksichtigt. En. „fish“ wird vom Hörer dem Kontext entsprechend als Einzahl oder Mehrzahl interpretiert, wozu an diesem Lexem keine formalen Mittel notwendig sind, wie bei cn. „唱“, das für alle Personen, Zeitformen usw. stehen kann. Auch bei morphologischen Alterationen wie Ablauten wird kein eigenes Morphem gezählt, da kein zusätzliches Sprachmaterial, das eine Bedeutung repräsentiert, vorhanden ist. Darüber hinaus wurden kumulative Morpheme wie es./pt. „-mos“, das für die erste Person Plural steht, als *ein* Morphem, wenn auch mit mehreren Bedeutungen (hier Person und Numerus betreffend), gezählt, da es nicht weiter zerlegt werden kann. Daher wurden auch Fugenelemente wie in „Wüstensturm“ (s. Bsp. (4.1) weiter unten) in der Morphemzählung berücksichtigt, auch wenn der Gehalt minimal ist und nur die Bedeutung „Bestandteil eines Kompositums“ vorliegt.

4.4.2.4 Definition von Wörtern als zählbare sprachliche Elemente

In Wood (1973a) werden mittlere Wortlängen zwischen 1,5 (Englisch) und 4,2 (Westgrönländisch) Silben gemessen. Für das Japanische sind, zumindest in bestimmten Kontexten und teilweise in Abhängigkeit von Sprecher/Geschlecht, Vergrößerungen der Wortlänge durch das Honorativpräfix お „o“ bekannt (s. z.B. Shibamoto, 1987: 28). Auch das sehr uneinheitlich lange Wort eignet sich demnach nicht uneingeschränkt zur vergleichenden Betrachtung des Tempos. Als Wörter wurden semantische Wörter gezählt, d.h. de. „Versicherungswesen“ wird ebenso wie en. „to put up with“ als ein Wort gezählt,

¹³⁹ Es muss hier nicht zwischen Morphemen und (Allo-) Morphen unterschieden werden, da die Anzahl im gesprochenen Text gleich ist.

¹⁴⁰ Bei Bloomfield (1933: 264) „Glosseme“, die (lexikalische) Morpheme und (grammatische) Tagmeme einschließen. Hier werden Morpheme als lexikalisch *und* grammatisch verstanden.

obwohl es sich im Englischen um mehrere orthografische Wörter handelt. Auch syntaktische Kriterien wie die Verschiebeprobe und phonologische wie die Wortbetonung wurden herangezogen, um Wörter zu segmentieren. In allen Zweifelsfällen wurden einsprachige Wörterbücher der entsprechenden Sprache konsultiert.¹⁴¹ Auch hierbei traten Problemfälle auf, wie im folgenden Transkriptionsbeispiel angesprochen wird.

4.4.3 Transkriptionsbeispiel

An dieser Stelle soll ein Beispielsatz einer deutschen Nachrichtensprecherin angeführt werden um auch die verwendeten Symbole zur Segmentierung zu veranschaulichen:

(4.1)

Orthografische Transkription

Der Staub stammt von Wüstenstürmen im völlig ausgetrockneten Hinterland Sydneys.

10 Wörter¹⁴²: Der} Staub} stammt} von} Wüstenstürmen} im} völlig} ausgetrockneten} Hinterland} Sydneys}

23 Morpheme: Der# Staub# stamm#t# von# Wüste#n#sturm#e#n# i#m# völl#ig# aus#ge#trock#ne#ten# Hinter#land# Sydney#s#

Breite ohrenphonetische Transkription (in Anlehnung an SAM-PA)

deStaubStamtfonvystnStymenimfölicausgetroknethintelantsIdnejs

Engere sonagrammgestützte Transkription

19 Silben (21 unter Berücksichtigung konsonantischer Silben), 64 Laute:

t6.Stäup.Stamt.fOn.vystn.Sty6.men.im.fö.lic.ʔaus.ge.tROk.netn.hIn.t6.lant.sId.neIs.

4.4.4 Pausenanalyse

Weder eine rein auditive noch eine optische Überprüfung im Oszillogramm alleine (s. Abb. 4.3 und 4.4) sind ausreichend genau, um stille Pausen als solche zu identifizieren. Im unten abgebildeten Ausschnitt, „[Steinmeier sei] als Fraktionsvorsitzender die richtige Besetzung“ wird das Prädikat „die richtige Besetzung“ durch eine kurze Pause (0,062 s)

¹⁴¹ Z.B. sind die jeweiligen Entsprechungen des Begriffs „Sicherheitsgurt“ in den einsprachigen Wörterbüchern von Michaelis (BP), 現代汉语词典 (CN), Duden (DE), Longman (EN), Priberam (EP) und RAE (ES) zu finden.

¹⁴² Obwohl der Begriff „Wüstensturm“ nicht im Duden gelistet ist, wird er hier aus semantischen Gründen als Wort gewertet.

vom Rest des Satzes wohl zur Hervorhebung abgetrennt. Diese Pause ist nur geringfügig länger als die Phase des totalen Verschlusses in „richtige“ (0,052 s). Bei der Analyse der Pausen wurde daher jede Unterbrechung im Sprachsignal auf ihren Kontext hin untersucht.

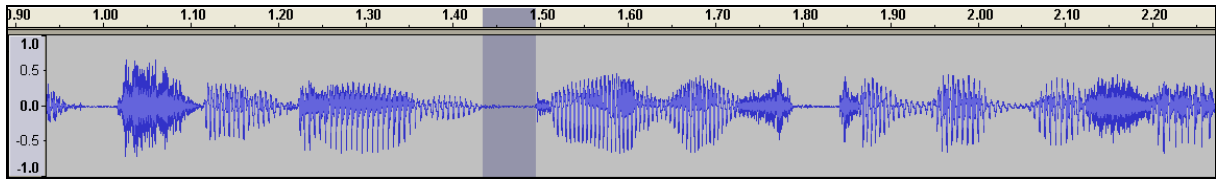


Abbildung 4.3: Kurze stille Pause (grau markierter Abschnitt) in „als Fraktionsvorsitzender die richtige Besetzung“

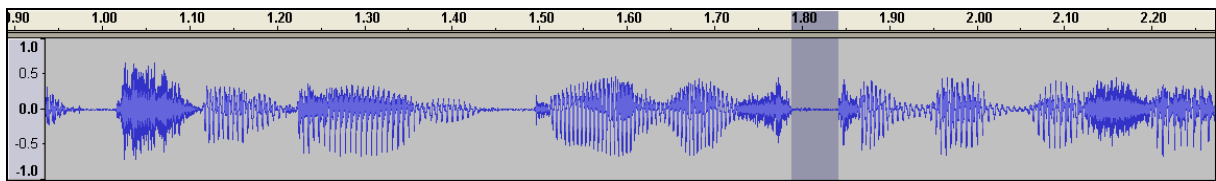


Abbildung 4.4: Phase des totalen Verschlusses (grau markierter Abschnitt) des [t] in „richtige“ in derselben Äußerung

Die Sprechzeit wurde vom Beginn des akustischen Sprachsignals bis zum Ende berechnet. Nur bei manchen Sprechern ist vor der Artikulation Atmung wahrnehmbar (das Einatmen vor der artikulatorischen Aktivität). Bei diesen Sprechern wäre es möglich gewesen, die Phase des ingressiven Luftstroms als erste Pause zu werten. Da dies jedoch nicht bei allen Sprechern möglich war und eine Pause vor der Artikulation nicht plausibel erscheint, da Pausen als Abschnitte der Stille innerhalb des Artikulierten definiert wurden, wurde auf die Berücksichtigung dieses Atemzuges verzichtet. Anders verhält es sich mit (Atem-)Pausen nach Beiträgen anderer Sprecher innerhalb einer Aufnahme. Sofern in den Nachrichtensendungen Beiträge eingespielt wurden, die nicht als Bestandteil einer Aufnahme in die Auswertung einfließen, wurden diese von der Sprechzeit abgezogen. Dabei wurde das Intervall vom Moment des Verstummens der Stimme des ausgewerteten Sprechers bis zum vollständigen Verstummen des nicht ausgewerteten Sprechers von der SZ abgezogen. Dadurch wird die Pause, die zwischen den beiden Redebeiträgen des ausgewerteten Sprechers stattfindet, mit berücksichtigt. Ein berechtigter Einwand hierbei ist, dass zwischen dem Ende des externen Beitrages und dem erneuten Einsetzen der Artikulation des ausgewerteten Sprechers eine durch die Strukturierung des Programms bedingte Verzögerung stattfinden kann, wenn z.B. der Videobeitrag länger andauert als die Berichterstattung des externen Reporters. In allen möglichen Fällen wurde dabei auf die Hintergrundgeräusche geachtet, die sich in der Regel unterscheiden: Im Tonstudio des auszuwertenden Nachrichtensprechers sind für gewöhnlich keine Hintergrundgeräusche vorhanden, wohingegen in Beiträgen anderer Reporter, v.a. Reporter vor Ort, Geräusche

vorhanden sind. Die Berücksichtigung der SZ des Nachrichtensprechers setzt in diesen Fällen mit dem Verstummen der Hintergrundgeräusche ein. In seltenen Fällen kann nicht davon ausgegangen werden, dass der Pausenbeginn des Nachrichtensprechers, d.h. der Moment, in dem die Hintergrundgeräusche des Beitrages verstummen und die Sprechpause des Nachrichtensprechers vor seiner Artikulation beginnt, exakt zu bestimmen ist. Diese geringen und seltenen Abweichungen mussten als Messungenauigkeit in Kauf genommen werden.

Von der SZ wurde die Summe aller Pausen abgezogen um die AZ zu berechnen. Die Summe der ausgezählten Laute und Silben wurde dann durch SZ und AZ geteilt; die Summe Wörter und Morpheme nur durch die SZ.¹⁴³ Auf diese Weise wurden globale Werte für alle Sprecher errechnet, die wiederum als Grundlage für die arithmetischen Mittelwerte und Streumaße der jeweiligen Einzelsprachen verwendet wurden.

Die statistischen Prüfwerte wurden mit einfaktoriellen Varianzanalysen und post-hoc-Tests (nach Scheffé) im Programm *SPSS* berechnet (abhängige Variablen: Laute/s (AG), Silben/s (AG), Laute/s (SG), Silben/s (SG), Morpheme/s (SG), Wörter/s (SG), Pausenanteil in Prozent, Pausendauer, Pausenhäufigkeit (=AZ/Anzahl Pausen); Faktor: Sprache). Geschlechtsspezifische Unterschiede wurden separat mit T-Tests untersucht. Allen statistischen Untersuchungen wurde ein Signifikanzniveau von 0,05 zugrunde gelegt.

4.5 Ergebnisse

Das untersuchte Sprachmaterial umfasst etwas mehr als 2,5 Stunden. Die einzelnen Sprachen sind zu annähernd gleichen Teilen enthalten, vgl. Abbildung 4.5. Tabelle 4.25 sind außerdem Anzahl und SZ der aufgenommenen Sprecher zu entnehmen. Das geringste Material liegt für BP vor (<19 min), das englische Material umfasst 34 min und ist am umfangreichsten (Mittelwert: 25,6 min).

¹⁴³ Morpheme und Wörter wurden als inhaltlich relevante sprachliche Elemente bezeichnet. Die artikulatorische Schnelligkeit wird deshalb nur anhand von Elementen gemessen, die sinnvoll artikulatorisch definier- und messbar sind.

Tabelle 4.25: Gerundete SZs (in s) der aufgenommenen Sprecher, deren arithmetische Mittelwerte als Basis für weitere Berechnungen dienen; fortan werden die Abkürzungen „w“ für weibliche Sprecher und „m“ für männliche Sprecher verwendet.

Sprecher	BP		CN		DE		EN		EP		ES	
	W	M	W	M	W	M	W	M	W	M	W	M
1	166	151	120	76	200	146	166	169	39	121	80	150
2	62	65	97	76	89	94	178	187	25	140	174	89
3	92	98	105	91	96	85	172	174	40	22	139	133
4	89	112	111	184	96	194	175	171	178	45	180	37
5	147	98	195	189	103	95	142	193	89	112	136	162
6	46	45	112	96	93	85	171	161	39	94	165	107
7			84	99	96				33	121		
8									87	216		
Summe	602	569	824	811	773	699	1004	1055	530	871	874	678
SD	42,9	34,0	33,2	45,3	36,7	40,4	11,9	10,9	48,0	55,6	33,8	42,0

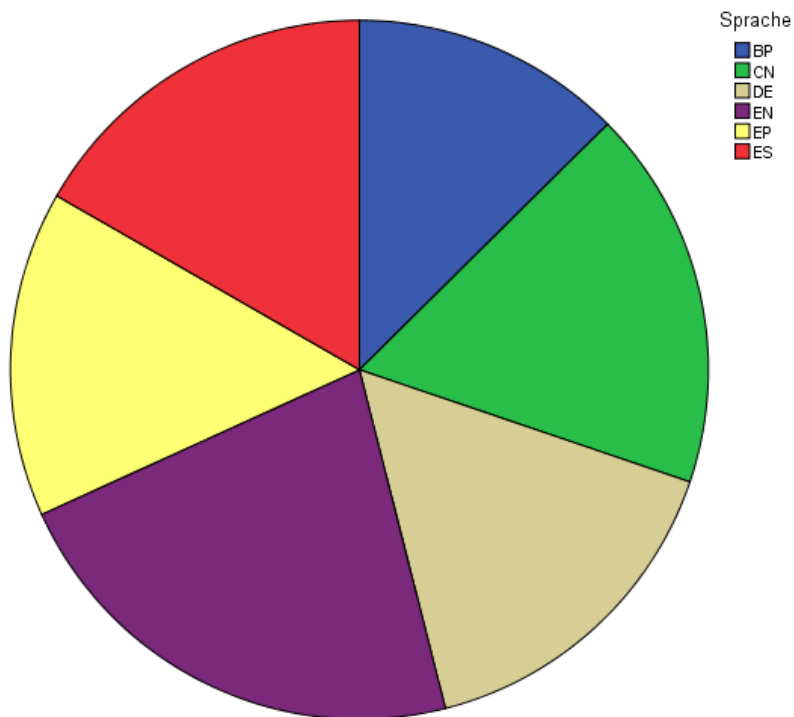


Abbildung 4.5: Anteil der SZ der untersuchten Sprachen am Gesamtkorpus

Anschließend werden die arithmetischen Mittelwerte der Sprechtempomesswerte, der Messwerte zur Untersuchung der Pausenstrukturen sowie die Silbenkomplexität und der Konsonantenanteil für die untersuchten Einzelsprachen präsentiert.¹⁴⁴

¹⁴⁴ Für einige Variablen werden keine geschlechtsspezifischen Werte angegeben, da der Faktor als irrelevant bezeichnet werden kann, z.B. bei der durchschnittlichen Anzahl an Silben/Wort, die zumindest in den hier untersuchten Sprachen ausschließlich von linguistischen Gegebenheiten abhängig sein dürfte.

4.5.1 Brasilianisch

Für Brasilianisches Portugiesisch wurde der Webcast „Globo Notícia para o G1“ untersucht, der zwei Mal täglich im Internet zur Verfügung gestellt wird und einen etwa vierminütigen Überblick über die wichtigsten aktuellen nationalen und internationalen Nachrichten gibt. Dieselben Sprecher moderieren die formellen und weniger formellen Nachrichtensendungen „Jornal Nacional“, „Jornal Hoje“ und „Jornal da Globo“, von denen „Jornal Nacional“ mit der Tagesschau um 20 Uhr vergleichbar ist. Das Material umfasst knapp 20 min SZ; darin wurden 16.370 artikulierte Laute und 7.026 Silben gezählt, was den in Tabelle 4.25 angegebenen Mittelwerten entspricht.¹⁴⁵ Der Konsonantenanteil (%K = Anteil konsonantischer Laute an allen artikulierte Lauten) beträgt 49,27%, was einem Konsonant/Vokal-Verhältnis von 1:1,08 entspricht; der Anteil von Halbvokalen und Halbkonsonanten ist verhältnismäßig hoch (9,64%). Weitere Parameter des Sprechtempos sind Informationsdichte (4,66 Morpheme/s) und Wortgeschwindigkeit (2,91 Wörter/s). 10,78% der SZ bestehen aus Pausen, davon werden 9,07% intern realisiert. Die mittlere Pausendauer beträgt 371 ms. Geschlechtsspezifische Unterschiede fallen nicht ins Gewicht, T-Tests auf Mittelwertgleichheit ergeben bei keiner Variable signifikante Unterschiede. Die Irrtumswahrscheinlichkeit liegt zwischen $p=0,73$ für Laute/Silbe, wobei $T_{(10)}=-2,001$ und $p=0,996$ für Morpheme/Silbe, wobei $T_{(10)}=-0,005$.

¹⁴⁵ Die in den Tabellen angegebenen Werte stellen die arithmetischen Mittelwerte aller Sprecher dar und weichen deshalb vom Quotienten der Gesamtwerte ab. Silben- und Lautgeschwindigkeiten werden nur in den Tabellen angegeben, um den fließenden Text nicht unnötig mit Zahlen zu überladen. Der Konsonantenanteil wird als absoluter Wert angegeben, mit „Pausendauer“ wird die durchschnittliche Länge stiller Pausen angegeben.

Tabelle 4.26: Sprechtempomesswerte - Ergebnisse für BP (arithmetische Mittelwerte und Standardabweichungen)

Variable		W	M	Gesamt	
				Mittel	SD
AG	Laute/s	16,06	15,85	15,95	0,78
	Silben/s	6,93	6,74	6,84	0,30
SG	Laute/s	14,43	13,97	14,20	1,03
	Silben/s	6,23	5,95	6,09	0,43
Wörter/s		2,95	2,87	2,91	0,21
Morpheme/s		4,77	4,54	4,66	0,42
Laute/Silbe		2,32	2,35	2,33	0,03
%K				49,27	
Pausen (%)		10,15	11,94	11,05	3,12
davon intern (%)		6,83	5,46	3,85	4,33
Pausendauer (ms)		370	373	371	96
Silben/art. Phrase		22,15	18,97	20,55	6,28

4.5.2 Chinesisch

Der staatliche Sender CCTV ist an der Verbreitung der chinesischen Normsprache 普通话 beteiligt.¹⁴⁶ Die Nachrichtensendung 新闻联播 (xīnwén liánbō, de. „Nachrichtensendung“) entspricht der deutschen „Tagesschau“ um 20 Uhr. Da in dieser Sendung nur wenige Sprecher die Nachrichten vorlesen, wurden daneben Aufnahmen weiterer Sendungen wie die Morgen-, Abend- und Nachtnachrichten (新闻 30 分, Xīnwén 30 fēn, de. „Nachrichten in 30 Min.“; 朝见天下, Cháojiàn Tiānxià, de. „Die Welt am Morgen“; 晚间新闻, Wǎnjiān Xīnwén, de. „Abendnachrichten“; 新闻 1+1, Xīnwén 1+1, de. „1+1 Nachrichten“; 中国新闻, Zhōngguó Xīnwén, de. „Nachrichten aus China“) untersucht. Die hier verwendeten Aufnahmen können beinahe ausnahmslos (vgl. Kap. 8) als formell bezeichnet werden und belaufen sich auf vierzehn Ausschnitte von je sieben Sprecherinnen und Sprechern mit einer Gesamtaufnahmezeit von über 27 min. Es wurden 25.676 Laute bzw. 8205 Silben erfasst, der Anteil konsonantischer Laute im gesprochenen Text liegt bei 62,53%¹⁴⁷ und auf einen Vokal treffen damit durchschnittlich 1,67 Konsonanten. Der Anteil von Gleitlauten ist mit 14,97% unter den behandelten Sprachen am höchsten. Die Morphemgeschwindigkeit liegt bei 5,34 Morphemen/s, die Anzahl der Wörter/s bei 2,38. Die Pausenzeit beträgt 11,3%, wovon 16,5% als interne Pausen realisiert werden, gemittelt

¹⁴⁶ Teilweise können in Nachrichtensendungen weniger formellen Formates umgangssprachliche Züge festgestellt werden, die in der Sendung 新闻联播 vollkommen fehlen, z.B. die Verwendung des Demonstrativpronomens 那个 mit der Aussprache [nèjkə].

¹⁴⁷ Aber vgl. hierzu Ergebnisse in 4.4.7.5.

dauert eine Pause 315 ms. Geschlechtsabhängige Unterschiede fallen außer bei der Anzahl interner Pausen nicht ins Gewicht, statistisch signifikante Unterschiede liegen für keinen Wert vor (der niedrigste Wert für p beträgt 0,13 und bezieht sich auf die Variable „Anteil interner Pausen“, wobei $T_{(12)}=-1,447$).

Tabelle 4.27: Sprechtempomesswerte - Ergebnisse für Chinesisch (arithmetische Mittelwerte und Standardabweichungen)

Variable		W	M	Gesamt	
				Mittel	SD
AG	Laute/s	18,71	18,21	18,46	0,87
	Silben/s	5,92	5,87	5,89	0,29
SG	Laute/s	16,7	16,31	16,56	0,78
	Silben/s	5,32	5,26	5,29	0,28
Wörter/s		2,37	2,39	2,38	0,17
Morpheme/s		5,37	5,3	5,34	0,32
Laute/Silbe		3,16	3,1	3,13	0,07
%K				53,04	
Pausen (%)		10,76	11,84	11,3	2,69
davon intern (%)		13,82	19,19	16,5	6,26
Pausendauer (ms)		312	320	315	64
Silben/art. Phrase		11,51	10,64	11,07	2,28

4.5.3 Deutsch

Für das Deutsche wurden die Nachrichtensendungen „Tagesschau“, „Tagesschau in 100 Sekunden“ und die Morgenausgabe (zwischen fünf und sieben Uhr) der „Tagesschau“ aufgenommen. Sieben Sprecherinnen und sechs Sprecher wurden mit einer Gesamtaufnahmedauer von etwa 25 min aufgenommen, insgesamt wurden 19.349 Laute bzw. 6.535 Silben gezählt. Da das Deutsche eine der Sprachen ist, in denen konsonantische Silben vorkommen können, wurde auch eine Silbenzählung durchgeführt, die nicht den vokalischen Kern zum Kriterium macht. Unter Berücksichtigung konsonantischer Silben ergeben sich dann 7061 Silben, was in Abschnitt 4.5.7.1 weiter kommentiert wird. Der geschlechtsspezifische Unterschied ist so gering (AG: $< 0,07$ Laute/s), dass er außer Acht gelassen werden kann, signifikante Unterschiede sind nicht festzustellen (niedrigster p -Wert: Morpheme/s: $T_{(11)}=1,969$; $p=0,75$). Der Anteil konsonantischer Laute liegt bei 62,12%, das Verhältnis von Vokalen zu Konsonanten lautet 1:1,64. Ein Anteil von 3,48% aller Laute wird als Halbvokal oder Halbkonsonant artikuliert. In Morphemen/s gemessen weist das Deutsche eine Geschwindigkeit von 3,82 auf, in Wörtern/s 2,16. Zur Pausenstruktur ist zu sagen, dass der Pausenanteil 13,67% beträgt und davon 10,5% interne

Pausen sind. Häsitationsphänomene wurden nicht beobachtet, stille Pausen sind durchschnittlich 600 ms lang.

Tabelle 4.28: Sprechtempomesswerte - Ergebnisse für Deutsch (arithmetische Mittelwerte und Standardabweichungen)

Variable		W	M	Gesamt	
				Mittel	SD
AG	Laute/s	15,66	15,56	15,59	0,60
	Silben/s	5,14	5,19	5,15	0,19
SG	Laute/s	13,7	13,43	13,48	0,54
	Silben/s	4,57	4,47	4,45	0,22
Wörter/s		2,21	2,19	2,2	0,09
Morpheme/s		3,91	3,71	3,82	0,21
Laute/Silbe		2,98	2,93	2,96	0,08
%K				62,12	
Pausen (%)		12,72	13,62	13,17	2,15
davon intern		10,46	10,53	10,5	10,40
Pausendauer (ms)		575	629	600	111
Silben/art. Phrase		19,6	20,62	20,07	4,17

Im Deutschen fällt auf, dass die phonetisch realisierte Lautkette für gewöhnlich streng mit der Form der kanonischen Lautkette übereinstimmt (s. Abb. 4.6).

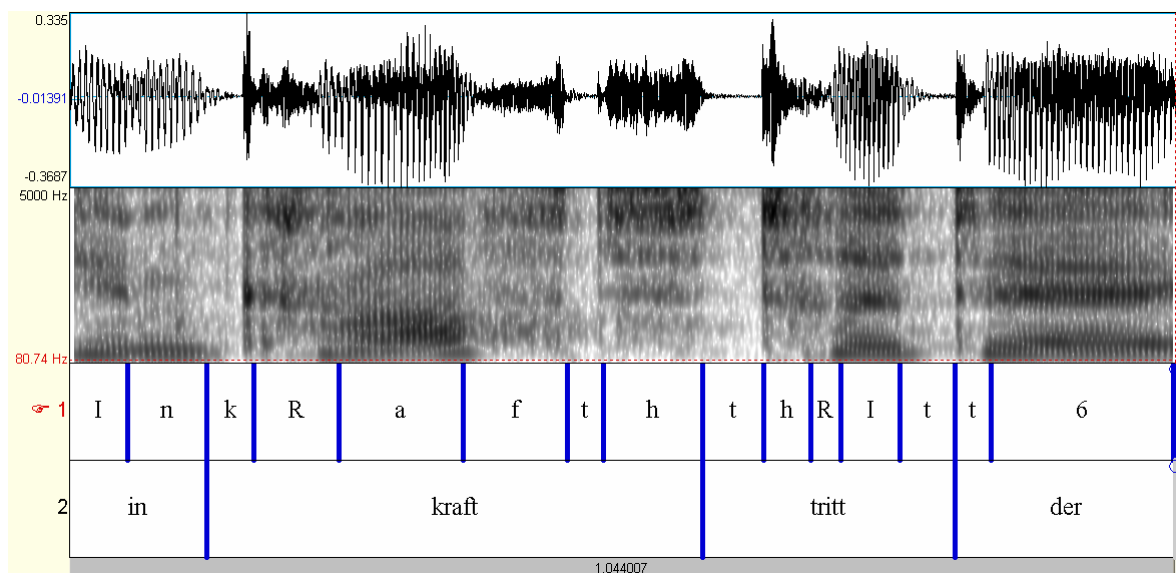


Abbildung 4.6: Komplexe Konsonantenhäufungen ohne Elision im Deutschen

4.5.4 Englisch

Englisch wurde anhand des stündlichen „News Bulletin“ der BBC News ausgewertet. Dazu wurden Sendungen von sechs Sprecherinnen und sechs Sprechern mit einer Aufnahmedauer von über 34 min aufgenommen. Es wurden 27.319 Laute und 9441 bzw. 9743 Silben (letztere ohne vokalischen Kern) gezählt. Der Anteil konsonantischer Laute liegt bei 58,63% und das Verhältnis von Vokalen zu Konsonanten beträgt 1:1,36. Halbvokale und Halbkonsonanten stellen 6,99% der artikulierten Laute dar. Morphem- bzw. Wortgeschwindigkeiten betragen 3,82 bzw. 2,91 Einheiten/s. Der Pausenanteil an der Sprechzeit liegt bei 12,37%, wovon 7,09% als interne Pausen realisiert werden. Neben diesen internen Pausen werden häufig Pausen am Phrasenrand eingefügt, hauptsächlich zur Hervorhebung bestimmter Satzteile. In seltenen Fällen werden Pausen innerhalb von Konsonantenhäufungen produziert (von einem Sprecher), durchschnittlich weisen Pausen eine Länge von 443 ms auf. Geschlechtsspezifische Unterschiede lassen sich bei allen gemessenen Werten kaum feststellen und sind nicht signifikant (niedrigster Wert für p: Morpheme/s, $T_{(10)}=-1,896$; $p=0,087$).

Tabelle 4.29: Sprechtempomesswerte - Ergebnisse für Englisch (arithmetische Mittelwerte und Standardabweichungen)

Variable		W	M	Gesamt	
				Mittel	SD
AG	Laute/s	15,47	15,88	15,58	0,62
	Silben/s	5,19	5,28	5,23	0,12
SG	Laute/s	13,5	13,98	13,74	0,60
	Silben/s	4,53	4,63	4,58	0,14
Wörter/s		2,79	2,88	2,83	0,08
Morpheme/s		3,78	3,85	3,82	0,06
Laute/Silbe		2,98	3,02	3,00	0,16
%K				58,63	
Pausen (%)		12,25	12,09	12,17	3,02
davon intern (%)		6,83	10,53	8,76	9,62
Pausendauer (ms)		462	423	443	54
Silben/art. Phrase		16,18	16,14	16,16	2,40

4.5.5 Portugiesisch

Da die Standard-Nachrichtensendung „Telejornal“ des Senders RTP im Normalfall von drei Nachrichtensprechern moderiert wird, mussten weitere Nachrichtensendungen, nämlich „Journal da tarde“ desselben Senders und „Journal2“, des Senders RTP2 untersucht werden. Insgesamt konnten damit pro Geschlecht acht Sprecher aufgenommen

werden. 17.243 Laute und 7230 Silben (7400 unter Berücksichtigung konsonantischer Silben) wurden in einer SZ von gut 23 min gezählt; es ergeben sich 2,38 Laute pro Silbe, der Konsonantenanteil liegt bei 51,69% und das Verhältnis von Vokalen zu Konsonanten beträgt 1:1,07. Ein Anteil von 7,65% der artikulierten Laute wird als Gleitlaute realisiert. Die Morphemgeschwindigkeit beträgt 4,65 Morpheme/s und die Wortgeschwindigkeit 2,8 Wörter/s. Der durchschnittliche Pausenanteil liegt bei 10,64%, wovon 20,77% interne Pausen sind; die durchschnittliche Dauer beträgt 396 ms. Im EP fallen Dehnungen und Hästitationsmittel auf, die unter 4.5.7.4 behandelt werden. Eine weitere Auffälligkeit im EP sind geschlechtsspezifische Unterschiede: Bei fast allen erhobenen Messwerten liegt ein deutlicherer Unterschied als in anderen Sprachen vor. Statistisch signifikant sind diese in den folgenden Fällen: Lautgeschwindigkeit (AG): $T_{(14)}=-2,386$; $p=0,032$; Silbengeschwindigkeit (AG): $T_{(14)}=-2,268$; $p=0,04$; und Pausenanteil: $T_{(14)}=-2,546$; $p=0,26$ (vgl. Abb. 4.5-4.7). Alle anderen Irrtumswahrscheinlichkeiten bezüglich der Gleichheit der Mittelwerte liegen deutlich über diesen Werten.

Tabelle 4.30: Sprechtempomesswerte - Ergebnisse für EP (arithmetische Mittelwerte und Standardabweichungen)

Variable		W	M	Gesamt	
				Mittel	SD
AG	Laute/s	13,79	15,08	14,44	1,24
	Silben/s	5,80	6,37	6,08	0,57
SG	Laute/s	12,62	12,75	12,68	1,17
	Silben/s	5,30	5,38	5,34	0,56
Wörter/s		2,76	2,80	2,78	0,29
Morpheme/s		4,57	4,65	4,61	0,49
Laute/Silbe		2,38	2,37	2,38	0,07
%K				31,49	
Pausen (%)		8,45	15,47	11,96	6,09
davon intern (%)		19,74	21,80	20,77	12,26
Pausendauer (ms)		314	455	385	161
Silben/art. Phrase		20,33	15,32	17,83	2,90

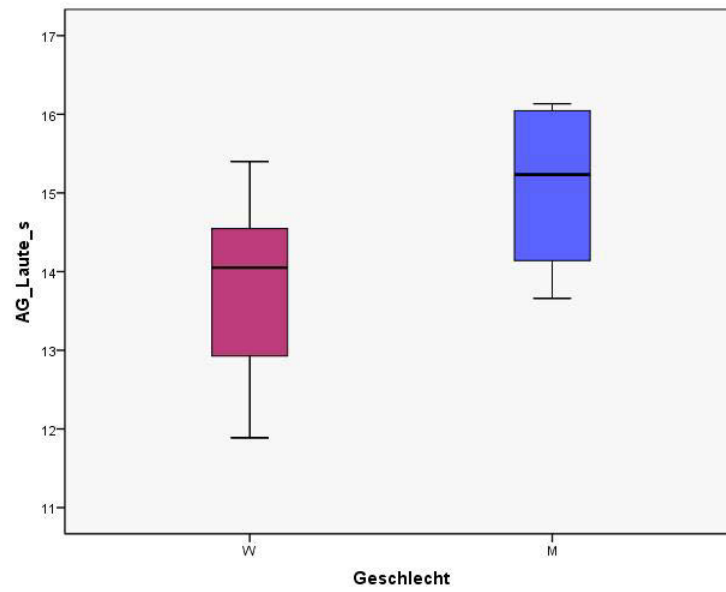


Abbildung 4.7: Geschlechtsspezifische Unterschiede im EP (AG in Lauten/s) in Quartildarstellung; 'w' = weibliche Sprecher, 'm' = männliche Sprecher

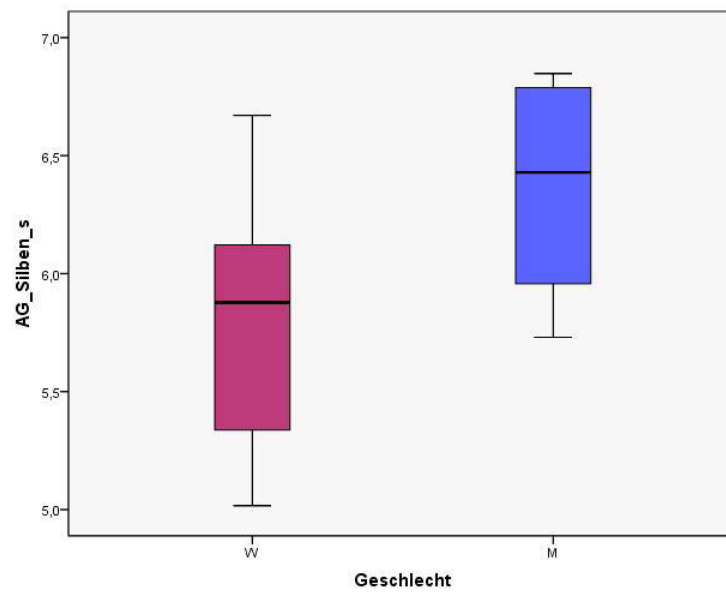


Abbildung 4.8: Geschlechtsspezifische Unterschiede im EP (AG in Silben/s)

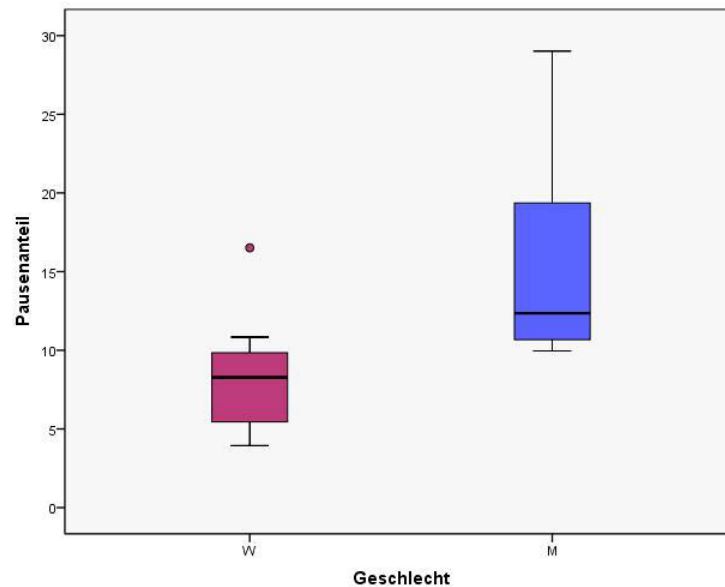


Abbildung 4.9: Geschlechtsspezifische Unterschiede im EP (Pausenanteil) mit Ausreißer (s.u.)

4.5.6 Spanisch

Als „Telediario en 4“ wird auf der Internetseite des spanischen Senders TVE der Nachrichtenüberblick zu Beginn der Nachrichtensendungen als Web-TV auf Abruf zur Verfügung gestellt. Wiederum wurden je sechs Nachrichtensprecherinnen und -sprecher aufgenommen und eine gesamte Aufnahmedauer von knapp 26 min erreicht. Da die Nachrichtensendungen in Spanien überwiegend von Sprecherinnen moderiert werden, mussten zwei Sportmoderatoren aufgenommen werden, die den Überblick über die Sportnachrichten in längeren Nachrichtensendungen präsentieren. Daneben wurde außerdem die Ansage eines Moderators des nächtlichen Informationsmagazins „La noche en 24 horas“ ausgewertet, das zwar nicht dem Stil der Nachrichtensendungen entspricht, jedoch als formell bezeichnet werden kann. Eine entsprechende Anzahl männlicher Sprecher zu erreichen wurde in diesem Fall als vorrangig angesehen. Das aufgenommene Sprachmaterial besteht aus 21.854 Lauten in 9.286 Silben. Der Konsonantenanteil beträgt 56,4% und das Verhältnis von Vokalen zu Konsonanten lautet 1:1,29. Der Anteil von Gleitlauten liegt bei 4,02%, wobei hier eine starke Präferenz von Halbkonsonanten (über 80% der Gleitlaute) gegenüber Halbvokalen besteht. Das Auszählen von Morphemen und Wörtern ergibt Geschwindigkeiten von 4,6 Morphemen/s und 3,03 Wörtern/s. Der Pausenanteil beträgt 16,74%, wovon 14,1% als interne Pausen realisiert werden. Pausen weisen eine mittlere Dauer von 575 ms auf. Im Spanischen kommt es nach Pausen oft zu einem vorgezogenen Stimmtoneinsatz (negative VOT) im hohen Bereich, aufgrund dessen im absoluten Anlaut vor dem eigentlichen Wortbeginn ein Laut wahrgenommen werden kann („n-retoma“, Abb. 4.10). Im absoluten Auslaut werden teilweise die letzten Silben

stimmlos produziert, also das Merkmal „Stimmlosigkeit“ vorgezogen. Obwohl die Texte in den spanischen Nachrichtensendungen vom Teleprompter abgelesen werden, belegen die häufigen Dehnungen und gefüllten Pausen einen umgangssprachlich geprägten Stil (s. 4.5.7.4). Auch im Spanischen sind signifikante Unterschiede zwischen den Geschlechtern nicht nachweisbar (kleinster Wert für p: Silbengeschwindigkeit (SG): $T_{(10)}=1,926$; $p=0,83$).

Tabelle 4.31: Sprechtempomesswerte - Ergebnisse für Spanisch (arithmetische Mittelwerte und Standardabweichungen)

Variable		W	M	Gesamt	
				Mittel	SD
AG	Laute/s	16,90	17,15	17,07	0,54
	Silben/s	7,25	7,25	7,25	0,24
SG	Laute/s	14,33	14,10	14,22	0,93
	Silben/s	6,12	5,96	6,04	0,40
Wörter/s		3,02	3,04	3,03	0,25
Morpheme/s		4,67	4,54	4,60	0,35
Laute/Silbe		2,34	2,37	2,35	0,03
%K				56,4	
Pausen (%)		15,69	17,79	16,74	5,03
davon intern (%)		18,15	10,05	14,10	13,53
Pausendauer (ms)		547	603	575	151
Silben/art. Phrase		21,38	20,24	20,81	6,68

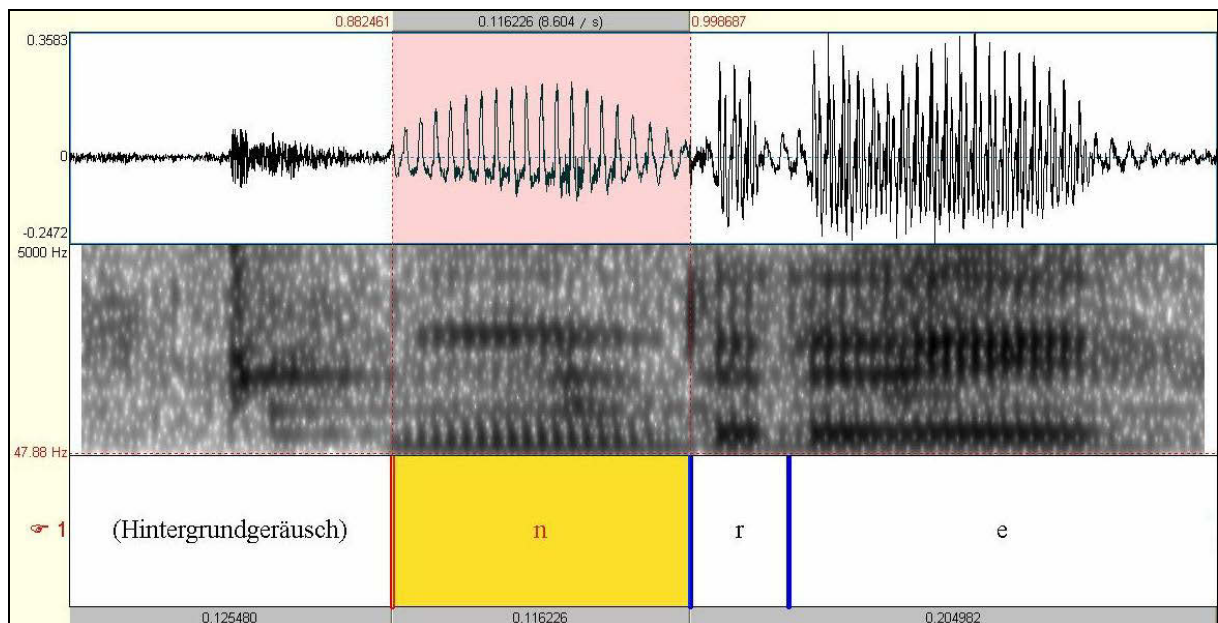


Abbildung 4.10: Negative VOT nach Pause in es. „retoma“ [nre'toma]

In spanischen Nachrichtensendungen können einige Reduktionen beobachtet werden (s. Abb. 4.11). Was die artikulatorische Genauigkeit anbelangt befinden sich damit die Sprachen Deutsch und Spanisch an den Extrempunkten eines Kontinuums.

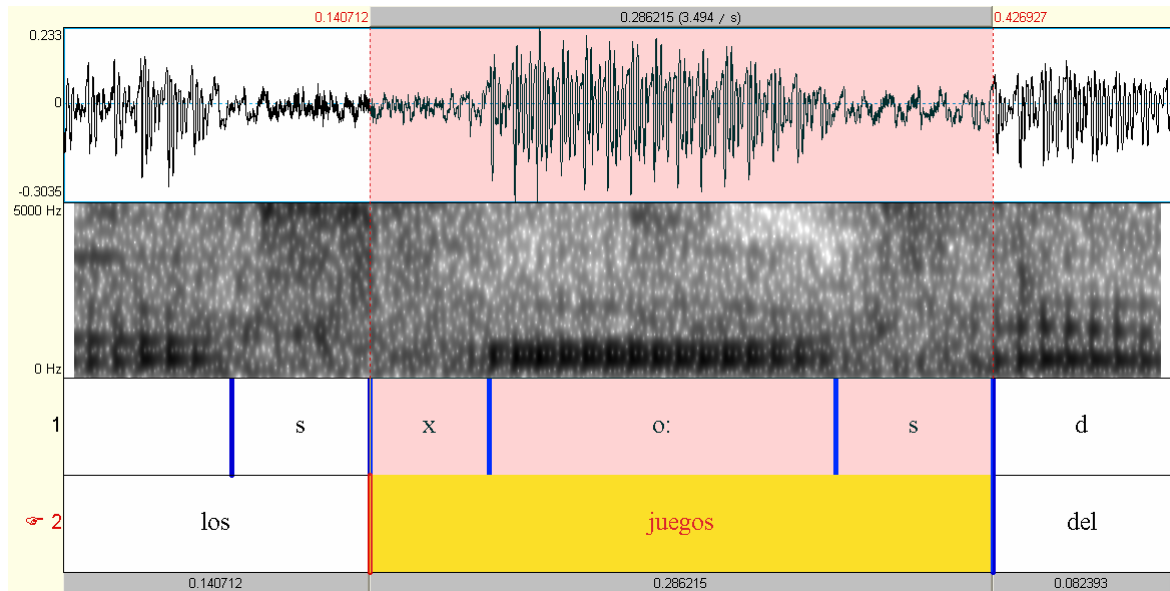


Abbildung 4.11: Stark reduzierte Aussprache von es. /'xweyos/

4.5.7 Vergleich

4.5.7.1 Tempo

Eine einfaktorielle Varianzanalyse¹⁴⁸ zeigt, dass sich die Sprachen in allen Parametern des Sprechtempos signifikant unterscheiden.¹⁴⁹ Die Sprache, die sich durch die höchste Lautgeschwindigkeit (AG) auszeichnet, ist das Chinesische, das sich signifikant von allen anderen Sprachen unterscheidet (ANOVA Post-Hoc Scheffé,¹⁵⁰ $p=0,004$ im Vergleich mit Spanisch; $p<0,001$ in allen anderen Vergleichen). Das Spanische weist die zweithöchste Lautgeschwindigkeit auf und unterscheidet sich mit Ausnahme des BP ($p=0,058$) signifikant von den anderen Sprachen ($p\leq 0,004$). Mit $p>0,999$ kann ein Unterschied

¹⁴⁸ In manchen Sprachen sind nicht alle Parameter des Sprechtempos normalverteilt, v.a. die Lautgeschwindigkeit nicht. Da für die Lautgeschwindigkeit sowohl der Kruskal-Wallis H-Test als auch der Median-Test wie die Varianzanalyse (hoch-)signifikante Unterschiede zwischen den Sprachen ergeben, wurden alle weiteren Tests auf Unterschiede in den Grundgesamtheiten mit der Varianzanalyse durchgeführt, die gegenüber der Verletzung der Normalverteilungsvoraussetzung relativ robust ist.

¹⁴⁹ Lautgeschwindigkeit (AG): $F_{(5,73)}=46,667$; $p<0,001$; Silbengeschwindigkeit (AG): $F_{(5,73)}=78,203$; $p<0,001$; Lautgeschwindigkeit (SG): $F_{(5,73)}=34,963$; $p<0,001$; Silbengeschwindigkeit (SG): $F_{(5,73)}=40,557$; $p<0,001$; Morphemgeschwindigkeit: $F_{(5,73)}=36,365$; $p<0,001$; Wortgeschwindigkeit: $F_{(5,73)}=32,917$; $p<0,001$.

¹⁵⁰ Alle weiteren Angaben zu signifikanten Gruppenunterschieden in diesem Kapitel basieren auf der Scheffé-Berechnung.

zwischen Deutsch und Englisch ausgeschlossen werden. EP unterscheidet sich nur von Deutsch und Englisch nicht signifikant. Damit können die sich überschneidenden Untergruppen wie folgt dargestellt werden, wobei homogene Untergruppen durch einen gemeinsamen Unter- bzw. Überstrich angezeigt werden und die arithmetischen Mittelwerte aufsteigend sortiert sind:

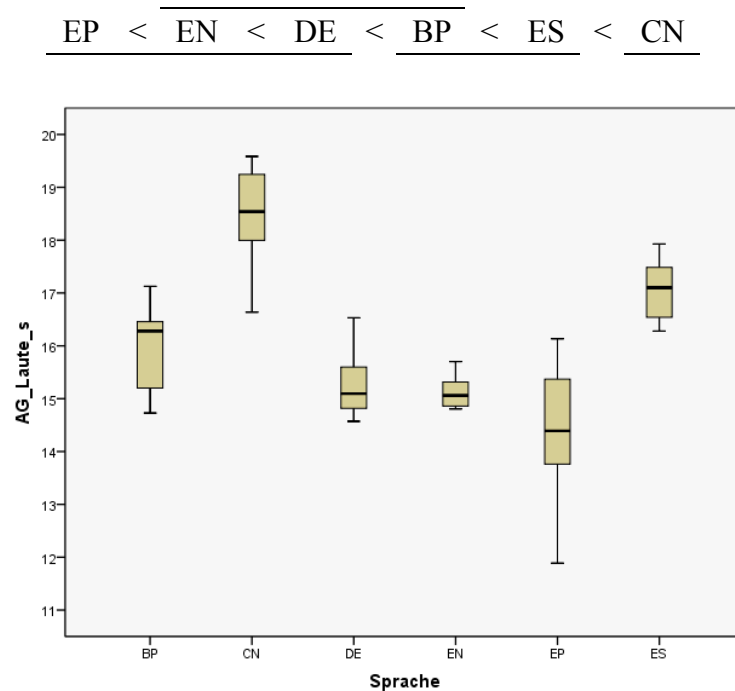


Abbildung 4.12: Lautgeschwindigkeiten (AG) der untersuchten Sprachen in Quartildarstellung

Bezüglich der Silbengeschwindigkeit (AG) stellt sich das Spanische, gefolgt vom BP, als schnellste Sprache heraus (kein signifikanter Unterschied). Deutsch und Englisch liegen wieder sehr nahe beieinander und unterscheiden sich nicht voneinander ($p=0,996$), jedoch beide von allen anderen Sprachen (jeweils $p<0,001$). Die Untergruppen nach Scheffé überschneiden sich nicht:

$$\underline{DE} < \underline{EN} < \underline{EP} < \underline{CN} < \underline{BP} < \underline{ES}$$

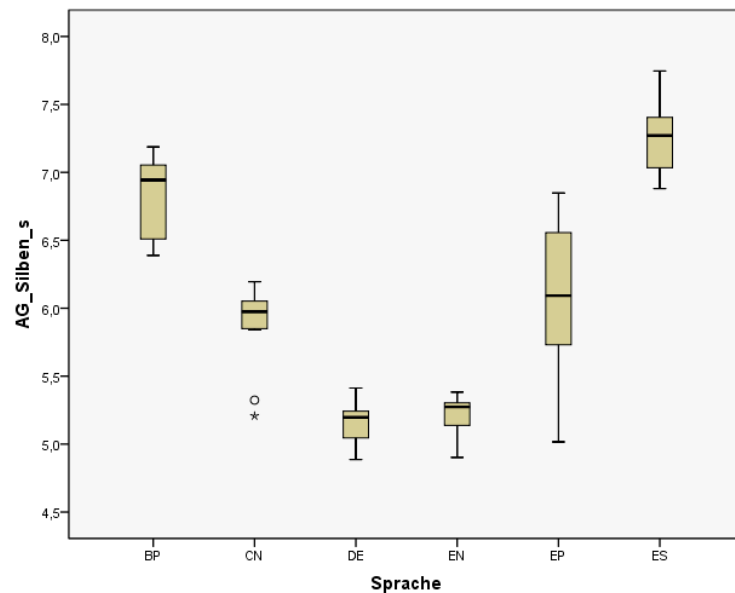


Abbildung 4.13: Silbengeschwindigkeiten (AG) in Quartildarstellung mit Ausreißern¹⁵¹ und Extremwerten¹⁵²

Werden Laut- und Silbengeschwindigkeit innerhalb der SZ erfasst, so ergibt sich optisch kein neues Bild, weswegen hier auf die Darstellung entsprechender Kastengrafiken verzichtet wird (s.u.). Die Varianzen fallen dabei etwas größer aus und für die Lautgeschwindigkeit ergeben sich abweichend nur drei Untergruppen (statt vier wie bei der AG):

$$\underline{EP < DE < EN} < BP < ES < \underline{CN}$$

Bei der Betrachtung der Wortgeschwindigkeit bilden sich zwei Untergruppen. Chinesisch und Deutsch ($p=0,236$ innerhalb der Gruppe) unterscheiden sich signifikant durch eine niedrigere Geschwindigkeit von allen anderen Sprachen ($p=0,099$ innerhalb der Gruppe):

$$\underline{DE < CN} < \underline{EP < EN} < BP < ES$$

¹⁵¹ Mit einem Kreis gekennzeichnete Ausreißer weisen einen Abstand von 1,5 bis 3 Kastenlängen (=mittlere 50% der Fälle) zum nächsten Quartil auf.

¹⁵² Mit einem Stern gekennzeichnete Extremwerte weisen einen Abstand von über drei Kastenlängen zum nächsten Quartil auf.

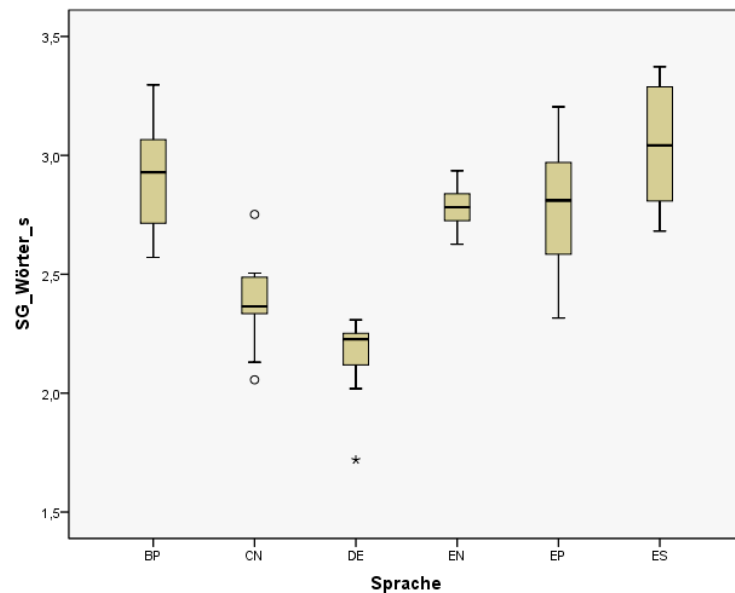


Abbildung 4.14: Wortgeschwindigkeiten der untersuchten Sprachen mit Median, mittleren Quartilen, Ausreißern und Extremwerten

Werden auch konsonantische Silben als eigenständige Silben gezählt („Lebensmittel“ als Viersilber), erhöhen sich die mittleren Silbengeschwindigkeiten (AG) im Deutschen auf 5,56; im Englischen auf 5,4 und im EP auf 6,19. Im Fall des Deutschen und des Englischen ergibt sich dann kein signifikanter Unterschied mehr zum Chinesischen (DE: $p=0,65$; EN: $p=0,126$), wobei alle anderen Unterschiede unverändert bleiben. Im Deutschen sind konsonantische Silben am häufigsten.

Insgesamt betrachtet zählt das Deutsche zu den langsameren Sprachen, auch das Englische weist in vielen Fällen Werte im unteren Bereich auf. Spanisch dagegen befindet sich immer unter den Sprachen mit hohen Sprechtempomesswerten. Chinesisch wiederum liefert eine überdurchschnittlich hohe Lautgeschwindigkeit, weist ansonsten jedoch Werte im mittleren Bereich auf. BP liegt im mittleren bis hohen Bereich; EP ist aufgrund unterschiedlicher Werte und hoher Schwankungen schwer zu kategorisieren. Abbildung 4.15 enthält auch die Morphemgeschwindigkeit, die in 4.5.7.3 besprochen wird. Deutlich erkennbar sind die niedrigeren Werte für Deutsch und Englisch:

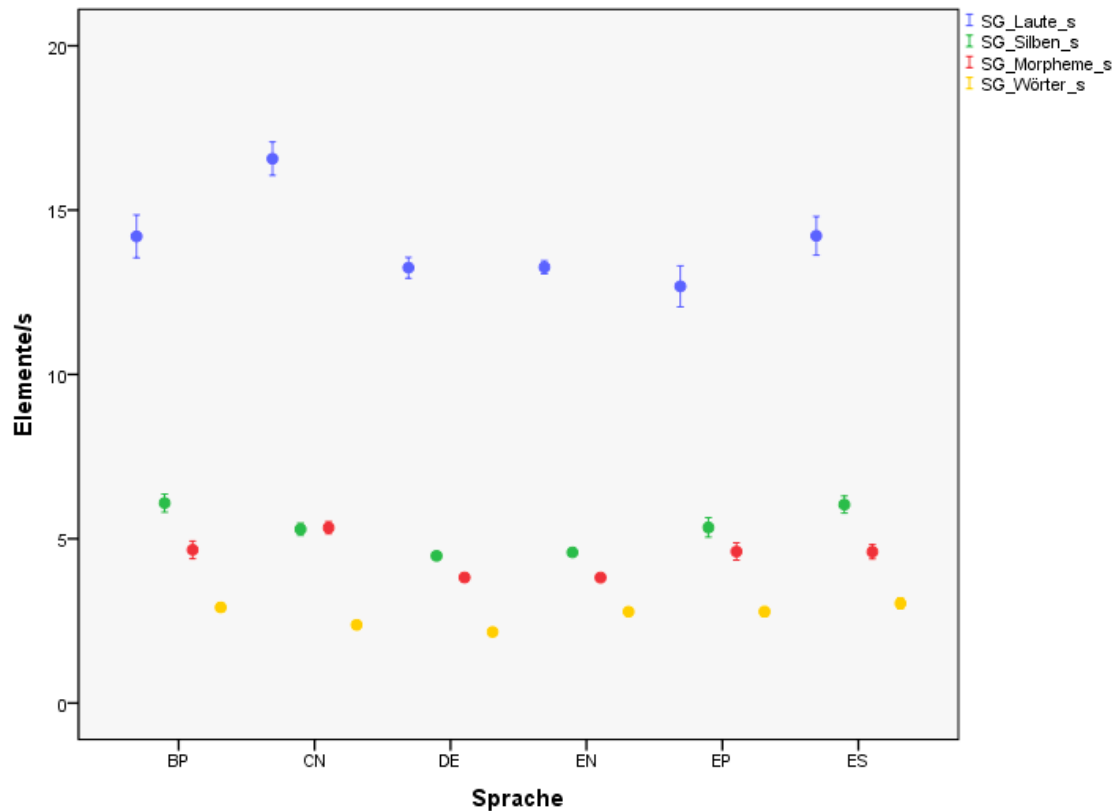


Abbildung 4.15: SGs der untersuchten Sprachen im Vergleich (arithmetische Mittelwerte mit 95%-Konfidenzintervallen der Mittelwerte)

Um die Frage zu beantworten, was die unterschiedlichen Geschwindigkeiten auslöst, wurden einige Korrelationen mit den für die Sprachen erhobenen Mittelwerten berechnet. AG und SG korrelieren stark positiv miteinander¹⁵³ (Lautgeschwindigkeit: $r=0,96^{**}$, Silbengeschwindigkeit: $r=0,979^{**}$).¹⁵⁴ Ein Hinweis auf kompensatorisches Verhalten der beiden Größen liegt damit nicht vor, ebenso scheint der Effekt sprachunabhängig zu sein.

Die Silbenkomplexität korreliert wie erwartet negativ mit der Silbengeschwindigkeit (AG: $r=-0,769$, SG: $r=-0,739$), d.h. die Silbendauer ist mit steigender Anzahl der Laute pro Silbe höher und damit die Silbengeschwindigkeit niedriger. Mit der Lautgeschwindigkeit korreliert die Silbenkomplexität nur mäßig und positiv (AG: $r=0,325$, SG: $r=0,426$), was auf eine Kompression innerhalb der Silbe bei steigender Lautzahl zurückgeführt werden kann.

¹⁵³ Alle auf 0,05-Niveau signifikanten Korrelationskoeffizienten werden mit ‘*’, auf 0,01-Niveau hochsignifikante mit ‘**’ angegeben; Korrelationskoeffizienten ohne Asterisk sind nicht signifikant, was teilweise mit der geringen Anzahl an Werten zusammenhängen kann.

¹⁵⁴ Eine Berücksichtigung des gesamten Sprachmaterials ohne Trennung nach Sprachen liefert vergleichbare Werte: Lautgeschwindigkeit: $r=0,894^{**}$, Silbengeschwindigkeit: $r=0,923^{**}$.

Da der Konsonantenanteil deutlich schwächer mit der Lautgeschwindigkeit (AG) korreliert ($r=0,17$), muss davon ausgegangen werden, dass eine höhere Lautgeschwindigkeit bei hoher Silbenkomplexität eher auf die Zunahme vokalischer Elemente (Di- und Triphthonge) zurückzuführen ist. Mit der Silbengeschwindigkeit (AG) korreliert der Konsonantenanteil stark negativ ($r=-0,645$). Die vermutete „bremsende Wirkung“ eines hohen Konsonantenanteils kann damit nur für die Silbengeschwindigkeit bestätigt werden.

Laut- und Silbengeschwindigkeit korrelieren mäßig miteinander (AG: $r=0,352$; SG: $r=0,292$), d.h. es handelt sich um weitgehend unabhängige Parameter.

Wider Erwarten korreliert der Anteil koronaler Konsonanten negativ mit der Lautgeschwindigkeit (AG): $r=-0,577$. Eine längere durchschnittliche Lautdauer bei höherem Anteil koronaler Konsonanten kann möglicherweise mit der Auslenkung zusammenhängen: Die in der Peripherie (ganz vorne im Mundraum) gebildeten Laute erfordern einen relativ weiten Weg, der von der Zunge bei der Artikulation zurückgelegt werden muss. Aufgrund der Einflüsse der lautlichen Umgebung auf die lokale Lautgeschwindigkeit muss aber davon ausgegangen werden, dass nicht die Beschaffenheit einzelner Segmente, sondern vielmehr das Zusammenspiel derselben ausschlaggebend ist.¹⁵⁵ Es ist nicht davon auszugehen, dass ein koronaler Konsonant in Umgebung hinterer Vokale die Lautgeschwindigkeit erhöhen kann.

Ein geringer Einfluss der segmentalen Beschaffenheit kann in Bezug auf den Anteil potentiell kürzerer Vokalphoneme vermutet werden. Dieser weist eine mäßige ($r=0,394$ mit der Lautgeschwindigkeit) bis mittlere Korrelation ($r=0,566$ mit der Silbengeschwindigkeit) mit den AGs auf.

Mit der Anzahl der Vokalphoneme, für die Manuel (1987, 1990) einen Einfluss auf den Grad der Überlappung nachweist, was wiederum potentiell Auswirkungen auf die Lautgeschwindigkeit haben kann, korrelieren alle Parameter des Sprechtempos negativ (darunter die Lautgeschwindigkeit (AG) mit $r=-0,639$; die Silbengeschwindigkeit (AG) mit $r=-0,525$). Ein kausaler Zusammenhang, etwa aufgrund einer längeren Mindestdauer von Vokalen zur eindeutigen Identifizierung des Phonems, wenn eine relativ hohe Anzahl von Vokalphonemen diskriminiert werden muss, ist rein spekulativ. Auffallend ist, zumindest im vorliegenden Sprachvergleich, dass in umfangreicheren Vokalsystemen tendenziell die

¹⁵⁵ Eine geringe Korrelation zwischen der artikulatorischen Organisation (Auslenkung, Geschwindigkeit und Beschleunigung der Artikulatoren) und der Laut-, Silben- und der *wahrgenommenen* Geschwindigkeit wurde auch in Tillmann/Pfützinger (2003) festgestellt.

Vokallänge zumindest als sekundäres Merkmal genutzt wird, was die durchschnittliche Lautdauer erhöht.

Die in absoluter Zeit gemessene Dauer artikulatorischer Phrasen korreliert negativ mit der AG (Lautgeschwindigkeit: $r=-0,760$; Silbengeschwindigkeit: $r=-0,321$). Der Zusammenhang deutet darauf hin, dass in der hier untersuchten Sprechsituation seltenere Pausen mit einer niedrigen AG einhergehen. Längere artikulatorische Phrasen können durch komplexere Sätze zustande kommen, deren niedriges Tempo als Hilfe für die problemlose Dekodierung verstanden werden kann. Voraussetzung hierfür ist eine weitgehend mit der syntaktischen Struktur übereinstimmende Pausenstruktur: Seltenere Pausen weisen dann auf längere Sätze, die mit niedrigerer AG einhergehen, hin. Eine Untersuchung des Zusammenhangs zwischen Tempo und syntaktischer Komplexität scheint aussichtsreich.

Die genannten und weitere Korrelationen werden in Tabelle 4.32 zusammengefasst:

Tabelle 4.32: Korrelationen zwischen Sprechtempomesswerten und segmentaler Beschaffenheit des Sprachmaterials (Sprachvergleich)

Tempomaß	Laute/Silbe	%K	Anteil koron. K	Anzahl V- phoneme	Anteil nicht- offener vorderer V
Laute/s (AG)	0,352	0,17	-0,577	-0,639	0,394
Silben/s (AG)	-0,769	-0,645	0,361	-0,525	0,566
Wörter/s	-0,808	-0,681	0,83	-0,373	0,312
Morpheme/s	-0,77	-0,409	-0,407	-0,752	0,278

Es fällt auf, dass bei den meisten Parametern des Sprechtempos die Werte für das Englische eine geringe Streubreite aufweisen. Dies kann als Hinweis darauf gedeutet werden, dass britische Nachrichtensendungen stärker normiert sind und daher einheitlicher gesprochen werden. Allerdings kann ausgeschlossen werden, dass genaue Vorgaben zum Sprechtempo gemacht werden, die verschiedene Aspekte desselben betreffen, da diese nur schwer von den Sprechern eingehalten werden könnten. Im Deutschen, das geringe Varianzen und ein niedriges Tempo aufweist, werden den Nachrichtensprechern vom Sender keine Vorgaben zum Sprechtempo gemacht (persönliche Mitteilung der Chefredaktion der ARD).¹⁵⁶ Auch die Nachrichtenredaktion des brasilianischen Senders Globo teilt mit, dass Nachrichtensprechern zum Sprechtempo keine Vorgaben gemacht

¹⁵⁶ Wachtel (2002: 130 ff.) hält fest, dass in Ratgebern zur Sprechweise im Fernsehjournalismus wenige und nur ungenaue Empfehlungen bezüglich des Sprechtempos gegeben werden, in denen tendenziell von (zu) hohem Tempo abgeraten wird.

werden.¹⁵⁷ Das portugiesische Sprachmaterial steht hier dem Englischen gegenüber und zeichnet sich im Allgemeinen durch eine auffallend große Streubreite aus, wozu sicherlich auch der deutliche geschlechtsspezifische Unterschied beiträgt.

4.5.7.2 Rhythmus

Die mittlere Anzahl der Laute/Silbe und die durchschnittliche Silbendauer weisen im Sprachvergleich, also bei Berechnung mit Hilfe der arithmetischen Mittelwerte, einen starken Zusammenhang auf ($r=0,757$). Der Korrelationskoeffizient ist statistisch nicht signifikant, allerdings weicht er nur geringfügig von dem statistisch hochsignifikanten Korrelationskoeffizienten ab, wenn alle Sprecher berücksichtigt werden ($r=0,680^{**}$). Da die Korrelation nicht vollständig ist, muss davon ausgegangen werden, dass längende (z.B. finale Dehnung) und komprimierende Effekte (z.B. Wortlängeneffekt) vorliegen. Diese müssen sehr unterschiedlich stark ausgeprägt sein, da innerhalb der einzelnen Sprachen die Zusammenhänge nicht gleich ausfallen (die überwiegend nicht signifikanten Werte für r lauten: BP: -0,250; CN: 0,343; DE: 0,219; EN: 0,610*; EP: 0,397; ES: 0,363). Dabei ist auch an die intrinsische Länge (z.B. bei gespannten Vokalen) zu denken. Überraschend sind hier die relativ niedrige Korrelation im Deutschen und der negative Zusammenhang im BP (je mehr Laute pro Silbe, desto kürzer die Silbe). Dabei ist zu beachten, dass eine hohe Silbenkomplexität im BP häufig auf einen komplexen Vokalkern zurückzuführen ist. Da das BP kaum Konsonantenhäufungen, jedoch zahlreiche Vokalbrechungen zulässt, muss davon ausgegangen werden, dass diese (z.B. Diphthongierung eines Nasalvokals) kürzer ausfallen als Monophthonge, die wiederum unter Akzent stark gedehnt werden können.

Bezüglich der Anzahl der Laute pro Silbe sind deutliche Unterschiede in den Standardabweichungen zu erkennen:¹⁵⁸ ES (0,029) < BP (0,031) < EN (0,050) < CN (0,071) < EP (0,074) < DE (0,090). Die Einheitlichkeit der Silben in Bezug auf ihre durchschnittliche Segmentzahl entspricht damit in etwa den oben dargestellten phonotaktischen Gegebenheiten.

Die Standardabweichungen der Silbendauern in den einzelnen Sprachen lauten:¹⁵⁹ CN (0,055) < ES (0,057) < BP (0,071) < DE (0,088) < EP (0,103) < EN (0,108). Der

¹⁵⁷ Leider haben die anderen Sender nicht auf entsprechende Anfragen geantwortet.

¹⁵⁸ Berechnung mit Durchschnittswerten für Sprecher.

¹⁵⁹ Diesen Berechnungen liegt eine Stichprobe von zwei Sprechern pro Geschlecht und Sprache zugrunde, deren produzierte Silben als Grundgesamtheit dienen.

phonologischen Rhythmusklassenterminologie der vergangenen Jahre entsprechend würde dies die Sprachen Chinesisch und Spanisch als am weitesten vom akzentzählenden Rhythmus entfernt, die Sprachen BP und Deutsch als intermediäre Sprachen und die Sprachen EP und Englisch als nahe am akzentzählenden Prototyp erscheinen lassen. Eine derartige Interpretation wird hier nicht verfolgt.

Für den wahrgenommenen Rhythmus kann davon ausgegangen werden, dass Sprechtempowechsel einen besonderen Einfluss ausüben. Die langsameren Sprachen Deutsch und Englisch, die geringe Varianzen zwischen den Sprechern aufweisen, scheinen dem Höreindruck nach auch ein global gesehen konstantes Tempo aufzuweisen. In Abschnitt 4.5.7.4 wird auf Veränderungen der Geschwindigkeit im Spanischen und im EP eingegangen. Diese Veränderungen tragen möglicherweise zu einem besonderen wahrgenommenen Rhythmus für Hörer anderer Muttersprachen bei.

Für die wahrgenommene Geschwindigkeit sind die einzelsprachlichen Korrelationen zwischen Laut- und Silbengeschwindigkeit (AG) von Bedeutung. Erwartungsgemäß fallen diese in den Sprachen mit einfacheren Silbenstrukturen höher aus (BP: $r=0,963^{**}$; EP: $r=0,943^{**}$; ES: $r=0,929^{**}$) als in den Sprachen mit komplexeren Silben (CN: $r=0,891^{**}$; EN: $r=0,706^{*}$; DE: $r=0,629^{*}$). Das Chinesische zeigt hier deutliche Ähnlichkeiten mit den Sprachen, die einfachere Silben bevorzugen.

4.5.7.3 Informationsdichte und Form der Kodierung

Ein weiterer Parameter des Tempos ist die Anzahl der Morpheme/s, die als Informationsdichte bezeichnet wurde. Auch hier sind die Unterschiede statistisch signifikant ($F_{(5; 73)}=36,356$; $p<0,001$), das Chinesische weist die höchste Geschwindigkeit auf. Die homogenen Untergruppen überschneiden sich nicht und entsprechen den genetischen Verwandtschaften:

$$\underline{\text{EN}} < \underline{\text{DE}} < \underline{\text{ES}} < \underline{\text{EP}} < \underline{\text{BP}} < \underline{\text{CN}}$$

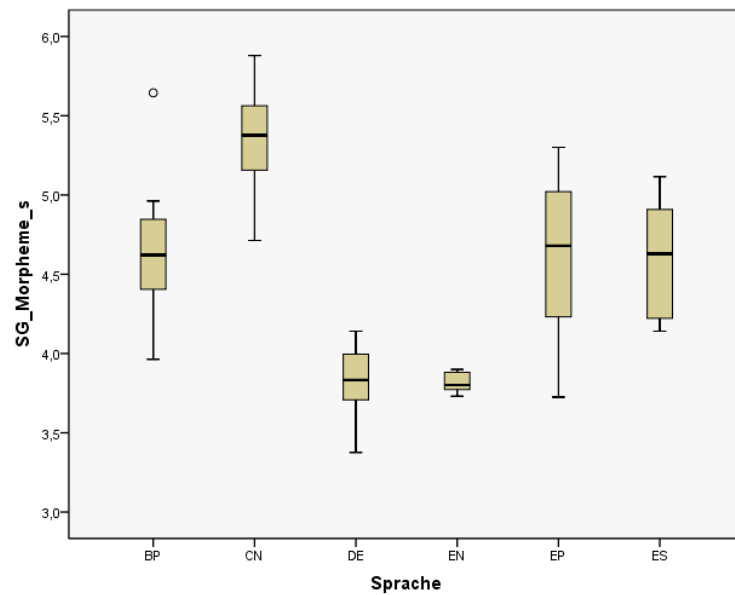


Abbildung 4.16: Morphemgeschwindigkeiten der untersuchten Sprachen mit Median, mittleren Quartilen und Ausreißern

Bei der Betrachtung inhaltlicher Konstituenten fällt das Englische durch etwas kürzere Wörter auf, die anderen Sprachen verwenden Wörter von etwa 1,9 bis 2,1 Silben Länge (s. u.). Eine Korrelation zwischen Silben- und Wortkomplexität (Fenk/Fenk-Oczlon/Fenk, 2005) kann nicht bestätigt werden ($r=0,03$; $p=0,995$):

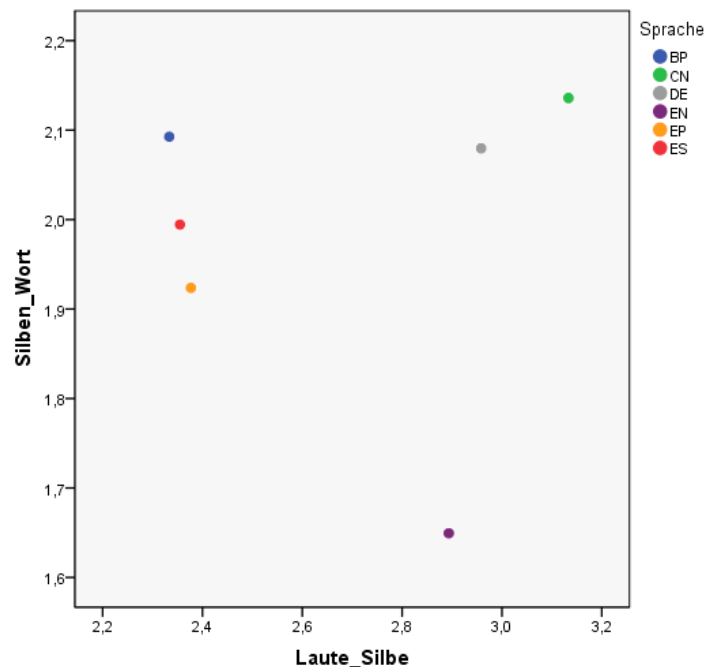


Abbildung 4.17: Zusammenhang zwischen Lauten/Silbe und Silben/Wort

Das kann z.T. am geringen Stichprobenumfang liegen. Gleichzeitig kann die Sprechsituation Zusammenhänge verdecken, wenn in manchen Sprachen in der

Nachrichtensprache besonders lange Wörter (z.B. Komposita) verwendet werden, was im Chinesischen und Deutschen der Fall zu sein scheint. Sollte eine hohe Silbenkomplexität tatsächlich mit kurzen Wörtern einhergehen, kann in Aussicht gestellt werden, dass eine hohe Laut- mit einer hohen Wort- und einer niedrigen Silbengeschwindigkeit zusammenhängt.

Im Chinesischen liegt die höchste Anzahl von Morphemen/Silbe (1,0), Silben/Wort (2,14) und damit auch von Morphemen/Wort (2,15) vor.

Obwohl das Chinesische kaum grammatische Morpheme kennt, zeichnet es sich unter den untersuchten Sprachen als die mit der höchsten Informationsdichte aus, sowohl was die Morphemgeschwindigkeit als auch die Morphemhäufigkeit (=dichte) betrifft.

Tabelle 4.33: Komplexität verschiedener sprachlicher Elemente der untersuchten Sprachen: arithmetische Mittelwerte und Standardabweichungen

Sprache	Laute/ Silbe		Morpheme/ Silbe		Silben/ Wort		Morpheme/ Wort		Laute/ Morphem	
	Mittel	SD	Mittel	SD	Mittel	SD	Mittel	SD	Mittel	SD
BP	2,33	0,031	0,77	0,035	2,09	0,074	1,60	0,084	3,05	0,134
CN	3,13	0,070	1,01	0,013	2,14	0,120	2,15	0,123	3,11	0,064
DE	2,96	0,092	0,85	0,049	2,08	0,152	1,77	0,133	3,47	0,190
EN	2,89	0,055	0,83	0,017	1,65	0,046	1,37	0,030	3,47	0,063
EP	2,38	0,074	0,86	0,046	1,92	0,080	1,66	0,063	2,76	0,121
ES	2,35	0,029	0,76	0,022	1,99	0,073	1,52	0,050	3,09	0,081

Diese Werte für Morpheme/Wort liegen unter den für Englisch (1,68) und Deutsch (1,92) bekannten.¹⁶⁰ Auf einen Wortlängeneffekt weist eine nur mäßige negative Korrelation zwischen Silben/Wort und Wörtern/s hin ($r=-0,357$): Da der Effekt („Je mehr Silben im Wort, desto weniger Wörter pro Sekunde“) nur relativ schwach nachgewiesen werden kann, muss davon ausgegangen werden, dass längere Wörter schneller produziert werden. Mit allen anderen Sprechtempoparametern korreliert die Wortlänge positiv ($r=0,347$ mit der Silbengeschwindigkeit; $r=0,524$ mit der Lautgeschwindigkeit; in beiden Fällen AG).

4.5.7.4 Pausenstruktur

Der Pausenanteil, definiert als prozentualer Anteil von Abschnitten der Stille während des Sprechens an der gesamten SZ, weist nur geringfügige Unterschiede zwischen den Sprachen auf, jedoch sind diese Unterschiede statistisch signifikant ($F_{(5; 73)}=3,821$;

¹⁶⁰ Haspelmath (2002: 5) gibt zu unterschiedlichen Morphemdichten (Morpheme/Wort) Werte u.a. aus Greenberg (1959) an.

$p=0,004$). Dabei unterscheidet sich Spanisch von BP ($p=0,027$) und Chinesisch ($p=0,022$). Die homogenen Untergruppen lauten:

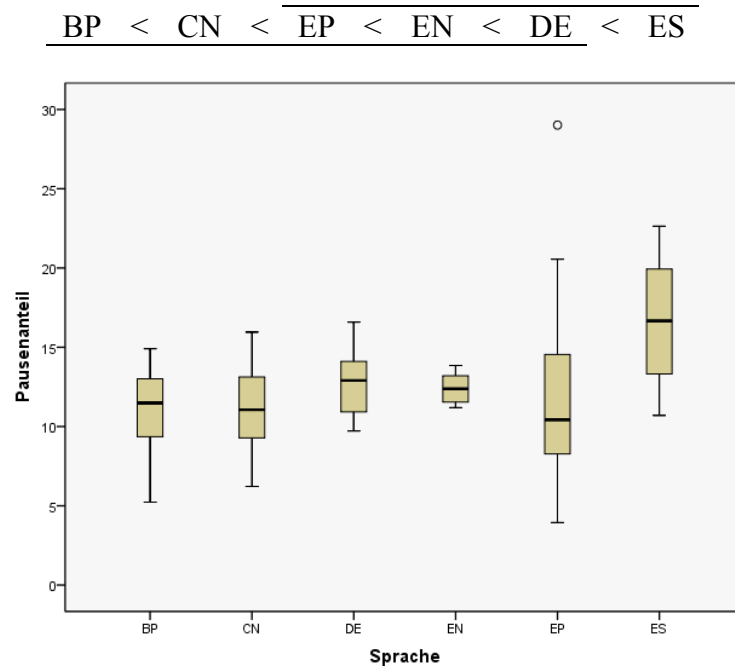


Abbildung 4.18: Pausenanteil in den untersuchten Sprachen (in %)

Der Anteil interner Pausen, definiert als prozentualer Anteil von phraseninternen Pausen an der Gesamtpausenzahl, weist ebenfalls signifikante Unterschiede auf ($F_{(5; 73)}=10,300$; $p<0,001$) und ist im EP höher als im BP ($p=0,001$), Deutschen ($p<0,001$) und Englischen ($p=0,003$); außerdem liegt er im Deutschen niedriger als im Chinesischen ($p=0,004$). Die Sprachen lassen sich, hier mit zusätzlicher Schattierung, wie folgt zu homogenen Untergruppen zusammenfassen:

DE < BP < EN < ES < CN < EP

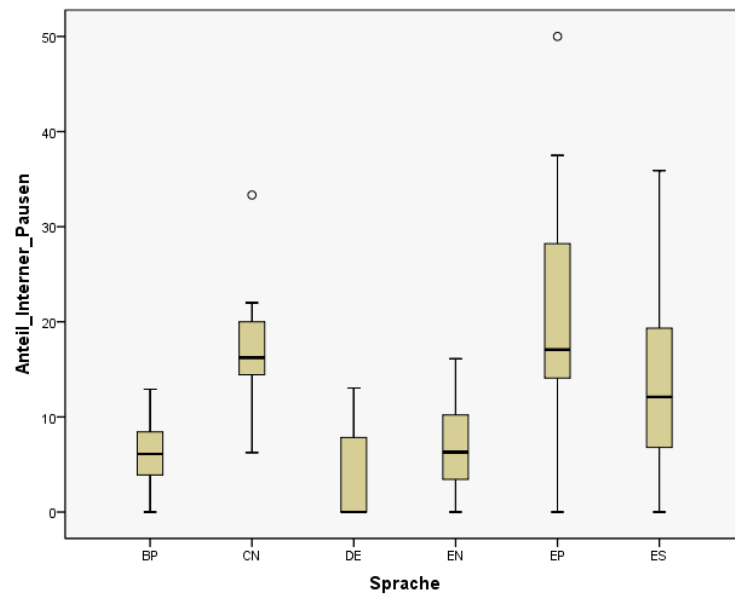


Abbildung 4.19: Anteil interner Pausen an allen Pausen (in %)

Hervorstechende Streuungsbreiten sind bei portugiesischen Sprechern in Bezug auf den Pausenanteil und bei portugiesischen Sprecherinnen bezüglich des Anteils interner Pausen zu sehen. Auch spanische Sprecherinnen zeigen hier starke Varianzen.

Diese Werte scheinen den Eindruck zu bekräftigen, dass in portugiesischen und teilweise in spanischen Nachrichtensendungen weniger stark standardisierte Sprechformen vorliegen. In beiden Sprachen werden Hässitationen produziert (damit sind hier Füllwörter und Dehnungen¹⁶¹ gemeint); im Spanischen sind Dehnungen so häufig, dass nur zwei Aufnahmen frei von ihnen sind (insgesamt wurden 64 Laute gedehnt, darunter Konsonanten und Vokale), und auch lexikalische Hässitationsphänomene sind relativ häufig (23 Füllwörter und -laute insgesamt¹⁶²). Als Vergleichswerte seien die Anzahl der Dehnungen in portugiesischen (15) und deutschen (2) Nachrichtensendungen angeführt. In den Aufnahmen der anderen Sprachen liegen keine Dehnungen vor.

Signifikante Unterschiede liegen auch bezüglich der mittleren Pausendauer ($F_{(5; 73)}=11,429$; $p<0,001$) und der Pausenhäufigkeit¹⁶³ vor ($F_{(5; 73)}=7,670$; $p<0,001$), s. folgende Darstellungen der Untergruppen und Kastengrafiken.

¹⁶¹ Unter einer Dehnung wird hier, perzeptiv betrachtet, die auffallend lange Dauer einzelner Segmente oder auch Silben verstanden, die nicht ausschließlich mit den Wort- und Satzakkzenten zusammenhängt, sondern einen verzögernden Eindruck hinterlässt.

¹⁶² Werden diese Phänomene als „gefüllte Pausen“ betrachtet und somit zu den Pausen gezählt, stellen sie einen Anteil von 4,7% der Pausen dar.

¹⁶³ Zur Berechnung der Pausenhäufigkeit (F_p , Frequenz von Pausen) kommen verschiedene Verfahren in Frage. Hier wird die in s gemessene absolute Dauer der artikulatorischen Phrase angegeben, weil dieses Maß

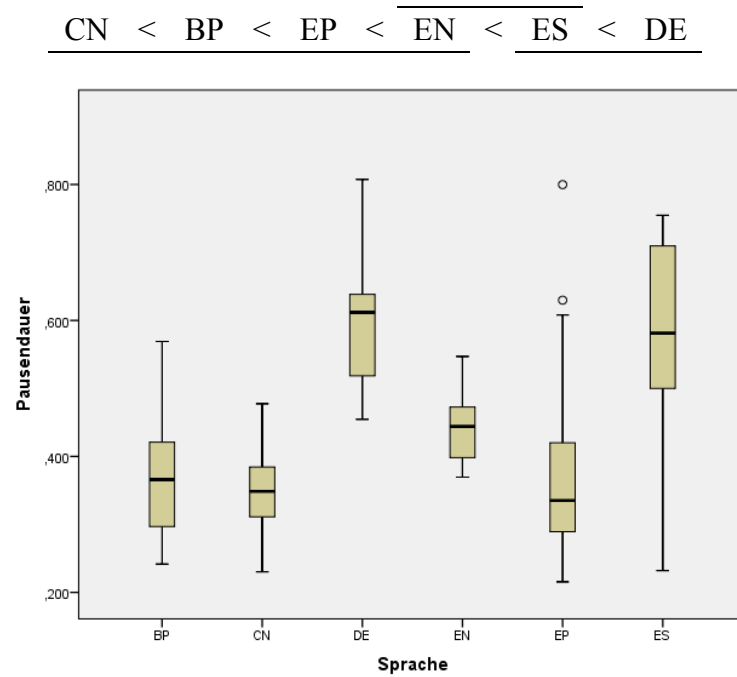


Abbildung 4.20: Durchschnittliche Pausendauer (in s)

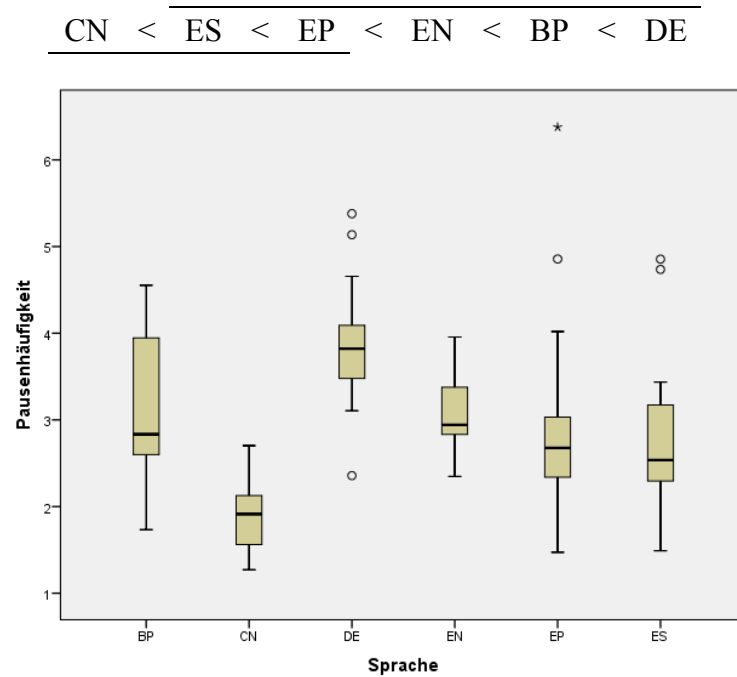


Abbildung 4.21: Durchschnittliche Pausenhäufigkeit (Abstand zwischen Pausen in s)

Auffallend sind v.a. die Pausen in deutschen Nachrichtensendungen, wo sie länger sind (durchschnittlich 600 ms gegenüber 348 ms im Chinesischen) und seltener produziert

bereits zuvor verwendet wurde. Es gibt den Abstand zwischen zwei Pausen (s/Pause) und damit die Umkehrung der Pausenhäufigkeit (Pausen/s) an. Als negativ formulierte Pausenhäufigkeit steht ein höherer Wert für seltenere Pausen.

werden (mit einem durchschnittlichen Abstand von 3,9 s zwischen zwei Pausen im Vergleich zu 1,9 s im Chinesischen). Der Pausenanteil ist in diesen beiden Sprachen dennoch vergleichbar (13% im Deutschen, 11% im Chinesischen). Zu den Mittelwerten und Standardabweichungen siehe Tabelle 4.33.

Tabelle 4.34: Pausenanteil, Anteil interner Pausen, Pausendauer und Dauer artikulatorischer Phrasen (jeweils arithmetische Mittelwerte und Standardabweichungen) der einzelnen Sprachen

Sprache	Pausen- anteil (%)		davon intern (%)		Pausen- dauer (s)		Phrasen- dauer (s)	
	Mittel	SD	Mittel	SD	Mittel	SD	Mittel	SD
BP	11,05	2,961	6,14	3,903	0,371	0,096	2,992	0,786
CN	11,13	2,723	16,87	6,313	0,348	0,064	1,883	0,390
DE	13,04	2,308	3,38	5,085	0,600	0,111	3,902	0,821
EN	12,37	0,936	7,09	5,092	0,443	0,054	3,088	0,442
EP	11,96	6,448	20,77	12,351	0,385	0,161	2,988	1,211
ES	16,74	3,807	14,10	9,915	0,575	0,152	2,865	1,016

Folgende Korrelationen geben Auskunft darüber, ob ein höherer Pausenanteil eher auf längere oder häufigere Pausen zurückzuführen ist: Im Gesamtkorpus (ohne Trennung nach Sprachen) korreliert der Pausenanteil positiv mit der Pausendauer ($r=0,616^{**}$), schwächer positiv mit der in s gemessenen absoluten Dauer artikulatorischer Phrasen ($r=0,368^{**}$). Diese Zusammenhänge scheinen sprachunabhängig zu sein, wie die vergleichbaren Korrelationskoeffizienten zeigen, die mit den Mittelwerten der einzelnen Sprachen berechnet werden. Sie lauten: $r=0,782$ (Zusammenhang von Pausenanteil und durchschnittlicher Pausendauer) und $r=0,223$ (Zusammenhang von Pausenanteil und Phrasenlänge). Es kann ein schwacher Zusammenhang zwischen selteneren Pausen und einem höheren Pausenanteil beobachtet werden, der sich durch die längere durchschnittliche Pausendauer seltenerer Pausen erklärt. Der letztgenannte Zusammenhang zwischen Pausendauer und Pausenabstand ist stark genug ausgeprägt ($r=0,392^{**}$ unter Berücksichtigung aller Sprecher ohne Trennung nach Sprachen bzw. $r=0,684$ im Sprachvergleich), dass festgehalten werden kann: In Sprachen, in denen längere Pausen produziert werden, sind diese zwar seltener, lösen aber dennoch einen höheren Pausenanteil aus. Die Gültigkeit muss selbstverständlich auf die hier untersuchten Sprachen und Redesituation eingeschränkt werden.

Der Pausenanteil hat keinen Einfluss auf die SG, wie eine mit den Mittelwerten der Einzelsprachen erstellte Korrelationsmatrix (Tab. 4.35) zeigt:

Tabelle 4.35: Korrelationen zwischen SGs und Pausen (Mittelwerte der Einzelsprachen)

Tempo-/Pausenmaß (SG)		Pausenanteil	Interne Pausen	Pausendauer	Phrasendauer
Laute/s	r	-0,173	0,220	-0,385	-0,814*
	p	0,743	0,675	0,452	0,049
Silben/s	r	0,225	0,335	-0,235	-0,406
	p	0,669	0,517	0,653	0,424
Wörter/s	r	0,352	0,240	-0,138	-0,162
	p	0,494	0,648	0,794	0,760
Morpheme/s	r	-0,205	0,677	-0,614	-0,876*
	p	0,697	0,140	0,195	0,022
Phrasendauer	r	-0,221	0,636	-0,611	
	p	0,674	0,174	0,198	
Pausendauer	r	0,782	-0,432		
	p	0,066	0,393		
Anteil interner Pausen	r	0,038			
	p	0,943			

Eine niedrigere Lautgeschwindigkeit (SG sowie AG, die stark mit der SG korreliert, s. 4.5.7.1) geht mit selteneren und gleichzeitig längeren Pausen einher. Damit muss ein potentieller Einfluss präpausaler Längung auf die Lautgeschwindigkeit (SG) in Form von Verringerung des Tempos als gering eingestuft werden, da häufige Längung (aufgrund häufiger Pausen) eben nicht mit niedriger Lautgeschwindigkeit einhergeht.

Anschließend werden Besonderheiten einzelner Sprachen vorgestellt. Im Chinesischen fällt zur Position interner Pausen auf, dass diese oft nach der Strukturpartikel 的 (de) stehen, die Besitz oder Zusammengehörigkeit ausdrückt. In der Aufnahme einer Sprecherin befinden sich fünf von acht internen Pausen in genau dieser Position, z.B.:

(4.2)

这场年会的 (0,077) 关键词
 DEM.PRON - ZEW¹⁶⁴ - jährliche Veranstaltung - STRUKTURPARTIKEL [Pause] - Motto
 „Das Motto der [Pause] diesjährigen Veranstaltung“

Gerade an dieser Position muss die Pause als nicht zu erwartende Pause bezeichnet werden. Da keine Dehnung der vorangehenden Strukturpartikel auffällt, kann eine verzögernde Funktion (z.B. Schwierigkeiten bei der Wortwahl) ausgeschlossen werden. Es muss sich um eine rhetorische Pause handeln, mit der der folgende, zum Verständnis des vorher gesagten notwendige Begriff betont wird.

¹⁶⁴ Zählseinheitswort für Veranstaltungen.

Neben der Pausenstruktur (gefüllte Pausen, Dehnungen, viele interne Pausen) zeigen auch intonatorische Besonderheiten, dass die Nachrichtensprache in Spanien von Phänomenen der Spontansprache geprägt ist. Ein Beispiel ist die intonatorische Beendigung einer Einheit durch Deklination, die syntaktisch unvollständig ist, siehe Abbildung 4.20.

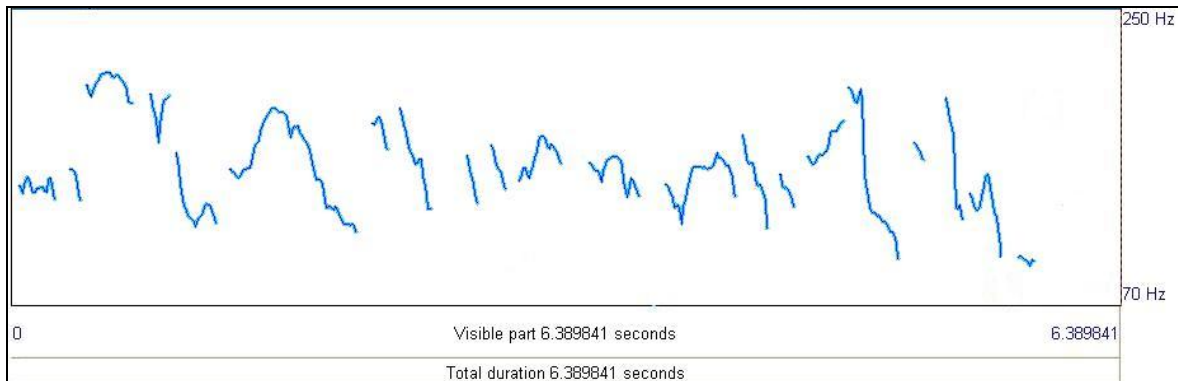


Abbildung 4.22: f0-Deklination in syntaktisch unvollständigem Satz bei spanischem Nachrichtensprecher. Text: „Y el primer diputado del PP en la asamblea de Madrid Alfonso Bosch que es el primer aforado formalmente imputado en el caso Gürtel“, de.: „Und der erste Abgeordnete der PP in der Landesregierung von Madrid Alfonso Bosch, der der Hauptbegünstigte im Fall Gürtel ist, der offiziell beschuldigt wird“.

Das Beispiel erinnert an die Überschrift in einer Zeitung und kann als Ankündigung des Themas der folgenden Meldung interpretiert werden. Daneben fallen z.B. bei einem der Sportmoderatoren Konstruktionsabbrüche und Füllwörter auf:¹⁶⁵

(4.3)

01 bueno ya hemos visto que todo empieza a girar de en torno

02 a esos: (0,191)

03 a esa decisión del: (0,130) que se va a tomar en copenhagen

De.: Nun, wir haben schon gesehen, dass alles anfängt sich darum zu drehen, um diese (Pl.) [Pause] diese (Sg.) Entscheidung des [Pause] die in Kopenhagen getroffen wird.

An späterer Stelle fällt eine wortinterne Pause mit anschließendem Durchhecheln auf:

¹⁶⁵ Folgende Transkription nach GAT-Konventionen ohne Akzente, ‘ : ’ = Längung, ‘ , ’ = leicht ansteigender f0-Verlauf, ‘ <accel> ’ = Beschleunigung, ‘ . ’ = abfallender f0-Verlauf, ‘ = ’ = schneller Anschluss (ohne „erwartete“ Pause), ‘ <all> ’ = schnell, ‘ <u> ’ = *Upstep*, Pausen werden mit Dauer in ms in Klammern angegeben.

(4.4)

01 los juegos del dos mil (0,372) dieciséis=los días están solo ?

< all >

De.: Die Spiele im Jahr zweitausend [Pause] sechzehn, es sind nur [schneller Anschluss, allegro]...

Im EP liegen einige Fälle vor, in denen anhand des f0-Verlaufs und der Pausenstruktur nicht eindeutig unterschieden werden kann, ob Frage- oder Aussagesätze bzw. hypo- oder parataktischer Satzbau vorliegt. Hier soll der Fall eines Sprechers vorgestellt werden, bei dem erwartete Pausen (zwischen Hauptsätzen) ausbleiben und unerwartete Pausen (zwischen oder innerhalb von Phrasen) produziert werden. Dazu findet ein starker Sprechtempowechsel zum Satzende hin statt:

(4.5)

01 o negócio envolve ainda a aquisição de mais duas empresas de capita:is

02 da privada holding que era até agora a dona (0,32)

03 do bpp.=a partir daqui: a ore: financial (0,44)

04 aguarda luz verde do banco de portugal.

< all >

De.: In das Geschäft verwickelt sind noch zwei Kapital [Dehnung] gesellschaften der privaten Holding-Gesellschaft, die bisher im Besitz der BPP war [schneller Anschluss] von jetzt an [Dehnung] wartet Orey Financial [Pause] auf grünes Licht der portugiesischen Zentralbank [allegro].

Im letzten Teil wird die Geschwindigkeit auf 8,46 Silben/s beschleunigt (von ursprünglich 5,77 Silben/s zu Beginn des Satzes). Am Ende des ersten Satzes (nach „do BPP“) wird zum folgenden Satz durchgehechelt, der ohne Pause angeschlossen wird, was bei mehreren Sprechern beobachtet werden kann. Weiterhin fällt auf, dass viele syntaktisch vollständige Einheiten intonatorisch nicht abgeschlossen werden, was Spannung erzeugt und ebenfalls einen direkten Übergang zum folgenden Satz generiert. In dem folgenden Auszug eines weiteren Sprechers wird an keiner Stelle eine Grenze intonatorisch, syntaktisch und durch eine Pause zugleich markiert:

und Nebensätzen zu, was bei Hörern mit anderer Muttersprache den Eindruck erhöhten Sprechtempos hinterlassen kann.

4.5.7.5 Chinesisch und Deutsch: Ein lokaler Vergleich

Um das Sprachmaterial auf Hinweise zu untersuchen, ob eine höhere Lautgeschwindigkeit auf kürzere Laute oder eine stärkere Überlappung zurückzuführen ist, wurden Stichproben aus der Sprache mit der höchsten (Chinesisch) und einer mit deutlich niedrigerer Lautgeschwindigkeit (Deutsch) genommen und die Dauer einzelner vergleichbarer Segmente sowie von Konsonantenhäufungen gemessen. Laute und Lautverbindungen, die sowohl im Chinesischen als auch im Deutschen nicht selten vorkommen, sind die aspirierte dentale Affrikate [ts^h]¹⁶⁷, der behauchte dentale oder alveolare Verschlusslaut [t^h] und der Sibilant [s]. Diese Laute und Lautverbindungen wurden in intervokalischer Position in der betonten Silbe des Wortes, die weder die erste noch die letzte in der artikulatorischen Phrase darstellt, gemessen. Es wurden folgende Mittelwerte berechnet:

Tabelle 4.36: Mittelwerte für die Konsonanten(-häufungen) [ts^h], [t^h] und [s] im Chinesischen und Deutschen

Messwert	Sprache	N	Mittel	SD
Dauer [ts ^h]	CN	11	0,0989	0,0342
	DE	11	0,1435	0,0471
Dauer [t ^h]	CN	16	0,0996	0,0232
	DE	12	0,1061	0,0222
Dauer [s]	CN	16	0,0952	0,0299
	DE	12	0,0952	0,0325

Aus den Mittelwerten geht hervor, dass ein größerer Unterschied nur bei [ts^h] auftritt. Ein T-Test bei unabhängigen Stichproben ergibt, dass zwischen der Dauer der chinesischen und deutschen Konsonantenhäufungen ein signifikanter Unterschied besteht ($T_{(20)}=-2,599$; $p=0,017$). Die Dauern von [t^h] und [s] weisen keine signifikanten Unterschiede auf ($T_{(25)}=0,238$; $p=0,814$ und $T_{(25)}=0,950$; $p=0,353$). Hier muss eingeräumt werden, dass die artikulatorische Beschaffenheit des Sibilanten vermutlich größere Unterschiede aufweist als dies bei [t^h] der Fall ist.¹⁶⁸ Wenn einzelne intervokalische Konsonanten in der Sprache,

¹⁶⁷ Zwar wird diese Affrikate im Deutschen als nicht aspiriert beschrieben, in betonter Stellung jedoch ist die Lautfolge nicht mit der unaspirierten Affrikate des Chinesischen [ts] vergleichbar.

¹⁶⁸ Die artikulatorische Beschaffenheit von /s/ im Chinesischen scheint nicht einheitlich zu sein, s. Lee (1999) zur Beschreibung als laminal alveolar, apikal dento-alveolar oder apikal dental, wohingegen /s/ im Deutschen in Kohler (1995: 161) als apikal alveolar beschrieben wird. Wie für Englisch bekannt ist, ist /s/ sprecherabhängig laminal oder apikal (Ladefoged/Maddieson, 1996: 146) und lässt einen relativ großen Spielraum im Artikulationsort zu.

die eine höhere Zahl an Lauten/s aufweist, nicht kürzer sind als in einer langsameren Sprache, gleichzeitig Konsonantenhäufungen jedoch signifikant kürzer sind, kann in Aussicht gestellt werden, dass in diesem Sprachenpaar tatsächlich eine stärkere Überlappung der einzelnen Segmente ausschlaggebend für die höhere AG im Chinesischen ist. Dabei weist die Ähnlichkeit in der Dauer des aspirierten Verschlusslautes darauf hin, dass hier keine unterschiedlichen Strategien vorliegen und die Artikulation als komplexe Artikulation *eines* Lautes aufgefasst werden kann. Die Ergebnisse weisen ebenfalls darauf hin, dass auch die chinesischen Konsonantencluster als *eine* komplexe Geste beschrieben werden können, da sie sich in ihrer Dauer nicht von einfachen Konsonanten unterscheiden (T-Test zum Vergleich der mittleren Dauern von [ts^h] und [s] im Chinesischen: $T_{(25)}=0,143$; $p=0,888$). Dies entspricht der traditionellen monosegmentalen Beschreibung der chinesischen Silbenanlaute.¹⁶⁹ Wird eine derartige Betrachtung artikulatorischer Segmente auf das vorliegende Sprachmaterial angewandt, so ergibt sich für die Lautgeschwindigkeit (AG) im Chinesischen ein Wert von 17,24 Lauten/s, der selbst dann für die hier behandelten Sprachen der höchste ist. Auch für die anderen Sprachen ergibt sich bei dieser Betrachtung ([t^h] als ein Laut gezählt) keine neue Rangfolge bezüglich der Lautgeschwindigkeit. Die monosegmentale Beschreibung behauchter stimmloser Verschlusslaute kann damit als ausreichend bezeichnet werden, um Unterschiede bezüglich der globalen Lautgeschwindigkeit zwischen Sprachen aufzudecken. Der Konsonantenanteil verringert sich damit im Chinesischen auf 50,28% und die Korrelation des Konsonantenanteils mit der Lautgeschwindigkeit (AG; Mittelwerte der Einzelsprachen) verringert sich auf $r=0,17$.

4.6 Zusammenfassung

Die Betrachtung verschiedener Parameter des Sprechtempos hat ergeben, dass das Spanische aufgrund seiner relativ hohen Laut-, Silben- und Wortgeschwindigkeiten (SG und AG) als schnell bezeichnet werden kann. Im Chinesischen wurden hohe Laut- und Morphemgeschwindigkeiten gemessen (SG und AG), ansonsten stellte es sich als Sprache mittlerer Geschwindigkeiten heraus. Langsame und zugleich ähnliche Sprachen bezüglich des Sprechtempos sind Englisch und Deutsch, die in diesem Vergleich niedrige Laut-,

¹⁶⁹ Diese Betrachtung geht auf das 601 von 陸法言 (Lù Fǎyán) herausgegebene Aussprachewörterbuch 切韻 (Qièyùn, de. etwa „Reim abschneiden“) zurück, in der allerdings auch die komplexen Auslaute als eine Einheit betrachtet werden. Die Beschreibung der Aussprache von Schriftzeichen beruht dabei auf der Zusammensetzung zweier Silben, deren Silbenkopf und –reim zusammengesetzt werden und ist bekannt als 反切 (fǎnqiè, de. „umgekehrt abschneiden“).

Silben- und Morphemgeschwindigkeiten aufweisen. Im Englischen liegt die Wortgeschwindigkeit höher als im Deutschen. Das BP zeigt eine Tendenz zu höherem Sprechtempo in allen Bereichen. Am schwierigsten ist das EP zu kategorisieren, weil es zugleich die niedrigste Lautgeschwindigkeit als auch mittlere Silben-, Wort- und Morphemgeschwindigkeiten bei starken Varianzen aufweist.

Rhythmusphonologische Unterschiede entsprechen nicht vollständig den Erwartungen. Zwar liegt die Standardabweichung der Anzahl der Laute/Silbe in den Sprachen mit einfachen Silben erwartungsgemäß deutlich niedriger. Allerdings fallen die Standardabweichungen der Silbendauern nicht dementsprechend aus. Möglicherweise haben Dehnungen, die auf diskursstrukturierende Phänomene zurückführbar sind, einen größeren Einfluss auf Abweichungen von der durchschnittlichen Silbendauer als die Silbenkomplexität.

Deutliche Unterschiede ergeben sich auch im Umgang mit Pausen. Der Pausenanteil, der im Allgemeinen bei etwas mehr als zehn Prozent liegt, fällt im Spanischen höher aus. Der Anteil interner Pausen ist sehr unterschiedlich und fällt im EP mehr als sechs Mal so hoch aus wie im Deutschen. Auch Pausendauer und Pausenhäufigkeit unterscheiden sich stark: Im Deutschen ist die durchschnittliche Pausendauer fast doppelt so lang wie im Chinesischen, dafür sind Pausen im Chinesischen etwa doppelt so häufig. Es kann auch gezeigt werden, dass Züge der Spontansprache nur im Spanischen und im EP zu beobachten sind, was auf unterschiedliche Konzeptionen von vorbereitetem Vorlesen in Nachrichtensendungen schließen lässt. In Kapitel 8 wird weiter auf derartige Unterschiede zwischen den Nachrichtensendungen eingegangen.

Auf der Suche nach erklärenden Variablen des Sprechtempos musste festgestellt werden, dass keine der beobachteten Korrelationen zwischen den Mittelwerten der Tempoparameter für Einzelsprachen und phonetisch-phonologischen Eigenschaften statistisch signifikant ist. Als Erklärung für die Silbengeschwindigkeit kann ein Einfluss der Silbenkomplexität bestätigt, weitere Faktoren jedoch nicht ausgeschlossen werden. Auf die Lautgeschwindigkeit hat eine hohe Silbenkomplexität dagegen beschleunigende Wirkung, ebenso ein hoher Anteil potentiell kürzerer Vokale. Für die Lautgeschwindigkeit ergibt sich außerdem ein negativer Zusammenhang mit dem Anteil koronaler Konsonanten sowie mit der Anzahl der Vokalphoneme.

Ein Vergleich der segmentellen Organisation im Chinesischen und im Deutschen auf lokaler Ebene gibt zunächst einen Hinweis darauf, dass tatsächlich der Grad der Überlappung eine höhere Lautgeschwindigkeit erklären kann, was im Einklang mit den in

4.2 beschriebenen phonotaktischen Beschränkungen im Chinesischen steht. Die Ergebnisse können auch so interpretiert werden, dass komplexe Konsonanten im Chinesischen als *ein* Laut artikuliert werden, wohingegen im Deutschen vergleichbare Bewegungen als eine Abfolge mehrerer Laute beschrieben werden können. Diese Erklärung verzichtet auf eine Berücksichtigung des Grades der Koartikulation.

Im folgenden Kapitel wird das Sprechtempo nicht vom produktiven, sondern vom perzeptiven Standpunkt aus untersucht.

5 Die Wahrnehmung des Sprechtempos

5.1 Einleitung

Bisher lag der Fokus bei der Untersuchung des Sprechtempos auf der Produktion. Anhand der Messwerte für die Anzahl sprachlicher Elemente pro Zeiteinheit wurde verglichen, wie schnell verschiedene Sprachen gesprochen werden. Darüber hinaus ist das Konzept *wahrgenommener* sprachlicher Elemente pro Zeiteinheit von Bedeutung. Hier werden folgende Fragen aufgegriffen: Klingen Sprachen gleich schnell? Was beeinflusst das wahrgenommene Sprechtempo? In diesem Kapitel wird nach einer Begriffsdefinition gezeigt, welche Unterschiede zwischen der Wahrnehmung des Sprechtempos in der Muttersprache und der in einer Fremdsprache bestehen können. Es werden außerdem das wahrgenommene Sprechtempo beeinflussende Faktoren vorgestellt. Anschließend behandelt Abschnitt 5.4 eine empirische Untersuchung zur Bedeutung von Silbenstrukturen für das wahrgenommene Tempo. Ziel dieser Untersuchung ist es, festzustellen, ob eine in der Muttersprache vorherrschende Präferenz für bestimmte Silbenstrukturen und damit der phonologische Typ der Muttersprache als Faktor des wahrgenommenen Tempos festgehalten werden kann. Die Zusammenfassung der Ergebnisse schließt den ersten, phonetisch-phonologisch orientierten Teil der Arbeit ab.

5.2 Definition

Neben den in 2.1.1 erwähnten Auffassungen von Sprechtempo existiert ein weiterer Begriff zur Definition der Schnelligkeit von Sprache. Die wahrgenommene Sprechgeschwindigkeit (wSG) muss nicht notwendigerweise mit in Größen wie Silben/s oder Laute/s gemessenen Geschwindigkeiten übereinstimmen. So wie Zeit als subjektiv erlebte Dimension verstanden wird (s. 2.1.2), die von physikalischer, messbarer Zeit abweichen kann, muss auch die wSG, da sie – wie Rhythmus – im Hörer entsteht, als ein potentiell von erhobenen Messwerten abweichendes Maß gelten. Objektiv gemessene Werte werden in Abgrenzung hierzu auch als LSR („laboratory measured speech rate“) bezeichnet. Es sind durchaus Regelmäßigkeiten festzustellen, wie dieses wahrgenommene Tempo beeinflusst wird.

Pfützinger (2001) weist nach, dass die lokale wSG im Deutschen (Spontansprache und vorgelesene Texte) einer Kombination aus den lokalen Laut- und Silbengeschwindigkeiten entspricht.¹⁷⁰ Wie sich zeigt, weist die Lautgeschwindigkeit einen Korrelationskoeffizienten von 0,836 und die Silbengeschwindigkeit einen von 0,803 mit der mittleren wSG (=gemittelte Einschätzung einzelner Stimuli durch Versuchsteilnehmer in Wahrnehmungsexperimenten) auf, die PLSR (=„Perzipierte Lokale Sprechrate“¹⁷¹) korreliert stärker mit der wSG ($r=0,959$). Die f_0 wird in dem Modell als Faktor der wSG abgelehnt, da sie nur für die Trainingsdurchgänge der Wahrnehmungsexperimente eine signifikante Verbesserung des Modells bewirkt. In Pfützinger (1999) stellt sich die Berücksichtigung der f_0 ohnehin als nicht ausschlaggebend heraus: r (zwischen wSG und PLSR) beträgt 0,9 ohne Berücksichtigung von f_0 und 0,91 mit ihrer Berücksichtigung. In Pfützinger (1998) wird festgehalten, dass die Silbengeschwindigkeit mit 54% einen stärkeren Einfluss auf die wSG hat als die Lautgeschwindigkeit. Seitdem unterstreicht Pfützinger, dass der Begriff „speech rate“ nicht verwendet werden sollte, wenn die objektiv gemessenen Größen Silben bzw. Laute/s gemeint sind, sondern der wSG vorbehalten bleiben sollte.

Grundsätzlich ist zu unterscheiden, ob gesprochene Sprache, wenn ihre Geschwindigkeit eingeschätzt wird, verstanden wird und damit Effekte, die auf Inhaltlichem oder Abweichungen von phonologischen Standardformen beruhen, nicht ausgeschlossen werden können, oder ob sie in Form eines akustischen Reizes präsentiert wird, dessen inhaltliche Seite von den Beurteilenden nicht berücksichtigt werden kann. Aus bisherigen Studien kann nicht eindeutig abgeleitet werden, dass in beiden Fällen dieselben Strategien bei der Einschätzung des Tempos angewandt werden. Als nicht verständliches Sprachmaterial gelten manipulierte und fremdsprachige Aufnahmen.

5.3 Muttersprache vs. Fremdsprache

Pfützinger (2001: 126) argumentiert gegen eine automatische Bestimmung der AG (Silben- bzw. Lautgeschwindigkeit) mit dem Argument, dass alleine der Mensch Sprache *verstehen* und Silben- und Lautgrenzen wahrnehmen kann und definiert damit die SG als das durch kognitive Prozesse der Dekodierung beeinflusste wahrgenommene Sprechtempo. Er fügt hinzu, dass die Verwendung von L1 und L2 von Bedeutung ist. Diese Betrachtung

¹⁷⁰ Dabei ist zu beachten, dass Silben- und Lautgeschwindigkeit auf der lokalen Ebene im Deutschen eine geringere Korrelation aufweisen ($r = 0,6$) als auf globaler Ebene (Pfützinger, 2001: 124).

¹⁷¹ Definiert als $s \cdot sr + p \cdot pr + c$, wobei sr = Silbengeschwindigkeit, pr = Lautgeschwindigkeit, c steht für die Rotation der Streuung der Daten; s , p und c werden mit der Lösung eines linearen Gleichungssystems nach der Methode der kleinsten Quadrate berechnet.

impliziert, dass die Einschätzung unbekannter Sprachen, mit deren Laut- und Silbenstrukturen Hörer nicht vertraut sind, ganz von den wahrgenommenen Laut- und Silbengrenzen abhängen. Dabei kann nicht davon ausgegangen werden, dass Hörer die lautliche Struktur fremder Sprachen „korrekt“, oder vielmehr dem phonologischen System der jeweiligen Sprache entsprechend, identifizieren. Im Fremdsprachenunterricht ist bei distanten Fremdsprachen (z.B. bei deutschen Lernern des Chinesischen) häufig das Phänomen zu beobachten, dass Lerner im Anfangsstadium nicht in der Lage sind, kurze Äußerungen (ohne deren Inhalt zu verstehen) in *Shadowing*-Aufgaben korrekt oder annähernd korrekt zu wiederholen.¹⁷² Das erklärt, warum die Einschätzung einer Sprache durch Hörer, die mit ihr nicht vertraut sind, von der der Muttersprachler dieser Sprache abweichen *kann* (s.u.).

Die Kenntnisse einer Fremdsprache haben einen Einfluss darauf, ob und wie sehr die Fremdsprache bezüglich des Tempos überschätzt wird. Schwab/Grosjean (2004) zeigen, dass Muttersprachler und schweizerische Lerner des Französischen langsame französischsprachige Aufnahmen etwa gleich einschätzen. Bei mittelschnellen Aufnahmen ist eine Überschätzung seitens der Fremdsprachenlerner zu beobachten, die bei schnelleren Aufnahmen stärker ausfällt. Da in der Untersuchung die 96 Fremdsprachenlerner Fragen zu den gehörten Texten beantworten, kann der Korrelationskoeffizient zwischen Verständnis (Anteil richtiger Antworten) und wSG (Werte auf einer freien Skala zur Beurteilung des Tempos) berechnet werden. Dieser ist negativ (-0,39) und signifikant, d.h. je weniger die Versuchsteilnehmer verstehen, desto höher schätzen sie die SG ein. Allerdings ist dieser Effekt nur für langsame Aufnahmen nachweisbar, bei zunehmendem Tempo geht er verloren (mittleres Tempo: $r=-0,21$, hohes Tempo: $r=0,05$). Die inhaltliche Verarbeitung der Aufnahmen ist sicherlich an der Einschätzung des Sprechtempos beteiligt, scheint aber einem Schwellenwert zu unterliegen, unter dem kein Zusammenhang mehr festgestellt werden kann. Die Fragen zu den schnellsten Texten werden nur zu 23,22% korrekt beantwortet. Dies kann als Hinweis darauf interpretiert werden, dass bei unverständlichen Sprachen eine neutralere Einschätzung stattfindet, da keine Beeinflussung der Einschätzung des Tempos durch kognitive Dekodierungsprozesse stattfindet.

¹⁷² Die *motor theory of speech perception* erklärt diesen Effekt mit dem Mechanismus der phonetischen Signalverarbeitung: Da an der Wahrnehmung von Sprachlauten die Erkennung der artikulatorischen Gesten zur Hervorbringung derselben beteiligt ist, Hörer jedoch mit der artikulatorischen Produktion dieser Laute (und Suprasegmentalia) nicht vertraut sind, können die Laute nicht präzise erkannt und folglich auch nicht nachgeahmt werden. Je nach Ähnlichkeitsgrad mit in der Muttersprache vorhandenen Lauten werden die Ziellaute dann ersetzt.

Den Os (1988: 78 ff.) stellt fest, dass, wenn italienische und niederländische Aufnahmen beurteilt werden, die wSG italienischer Versuchsteilnehmer am stärksten mit der Lautgeschwindigkeit korreliert. Die wSG niederländischer Versuchsteilnehmer dagegen korreliert bei niederländischen Stimuli am stärksten mit der in kanonischen Silben gemessenen Geschwindigkeit ($r=0,94$) und schwächer mit der in tatsächlich realisierten Silben Gemessenen ($r=0,85$); die Korrelation mit tatsächlich realisierten Lauten/s liegt dazwischen ($r=0,90$). Für die Beurteilung italienischer Stimuli ergibt sich kein einheitliches Bild, da die Korrelationskoeffizienten nur zwischen 0,92 und 0,96 schwanken und zudem starke Korrelationen (0,96 – 0,98) zwischen den einzelnen objektiv gemessenen Geschwindigkeiten bestehen. Da die Versuchsteilnehmer in beiden Fällen nicht mit der Fremdsprache vertraut sind, ist davon auszugehen, dass sie die kanonische Lautzahl nicht kennen. In einer weiteren Untersuchung wird die Tendenz gefunden, dass italienische Aufnahmen von niederländischen Muttersprachlern als etwas langsamer empfunden werden als niederländische Aufnahmen. Unbekannte Fremdsprachen müssen also nicht generell als schneller bewertet werden. Zu berechtigten Zweifeln an der Lautgeschwindigkeit als alleiniger Faktor der wSG gibt die fehlende Diskretheit von Segmenten Anlass, auch wenn das Vorhanden- oder Nicht-Vorhandensein einzelner Laute die wSG beeinflussen kann (s.u.). Ebenso ist die Silbe als alleinige Kategorie, auf die die wSG sich stützt, unwahrscheinlich, da nicht ersichtlich ist, warum ausschließlich Abfolgen von B-prosodischen¹⁷³ Elementen das wahrgenommene Tempo beschreiben sollen, wenn dieses global definiert wird und somit als ganzheitliches Phänomen einer Äußerung verstanden wird. Lokale Änderungen können durchaus von B- und C-prosodischen Elementen beeinflusst sein.

Den Os' Erklärung zur Beurteilung des Tempos fremder Sprachen erinnert an die Untersuchung von Osser/Peng (1964), in der in Aussicht gestellt wird, dass Japaner, wenn sie Englisch hören, *Laute* „zählen“, da sie nicht mit der Silbenstruktur des Englischen vertraut sind. Möglicherweise sind sich diese Hörer ihrer Strategie nicht bewusst und meinen, Silben zu zählen.¹⁷⁴ Zumindest für japanische Hörer ist unter Berücksichtigung

¹⁷³ In der Terminologie von Tillmann/Mansell (1980) bezeichnet die B-Prosodie Veränderungen von Klangeigenschaften mit genau der Geschwindigkeit, die eine Wahrnehmung rhythmisch gegliederter Klangereignisse zur Folge hat, im Gegensatz zur A-Prosodie, die eine langsamere Geschwindigkeit und folglich eine wahrgenommene kontinuierliche Veränderung bezeichnet, und der C-Prosodie, die eine hohe Geschwindigkeit und folglich ein konstantes, nicht in Teilereignisse zerlegbares wahrgenommenes Signal benennt.

¹⁷⁴ Ähnliche Effekte sind für Muttersprachler des Chinesischen festzustellen, die en. „bus“ als Zweisilber bezeichnen, ebenso wie das aus dem Englischen entlehnte 巴士, „bashi“, nach Duanmus Analyse zweisilbig [baʃz].

der Ergebnisse von Parlato-Oliveira et al. (2010)¹⁷⁵ plausibel, dass Laute mit Silben gleichgesetzt werden und in Konsonantenhäufungen „illusorische“ Vokale eingefügt werden, die die Laute des Englischen für Japaner zu wahrgenommenen Silben transformieren (vgl. Pfitzinger/Tamashima, 2006). Dabei scheinen die phonologisch zugrunde liegenden phonotaktischen Verhältnisse in der Muttersprache der Hörer ausschlaggebend zu sein: Im Japanischen kann es zwar in gesprochener Sprache durchaus zu Konsonantenhäufungen kommen (つくえ, „tsukue“, de. „Schreibtisch“, [tskwe]), diese erreichen jedoch in der Regel nicht die Komplexität englischer Konsonantenhäufungen und die einzelnen Silben des Wortes werden in isolierter Form nicht ohne vokalischen Kern gesprochen ([tsu.ku.e]). Außerdem gibt es Beweise dafür, dass diese auch von Japanern selbst in der Wahrnehmung durch illusorische Vokale ergänzt werden (Dupoux et al., 1999).

Pfitzinger/Tamashima (2006) erklären das Ergebnis von den Os (1985)¹⁷⁶ wie folgt: Da Italienisch durchschnittlich weniger Laute pro Silbe aufweist und das Sprachmaterial beider Sprachen in der Untersuchung mit vergleichbaren Silbengeschwindigkeiten produziert wurde, leuchtet ein, dass Hörer beider Sprachen das italienische Sprachmaterial als langsamer wahrnehmen, da weniger Laute/s als im Niederländischen produziert werden. Die Autoren gehen davon aus, dass Hörer nicht produzierte Laute in Fremdsprachen wahrnehmen, die das Material phonotaktisch mit der Muttersprache kompatibel machen. Allerdings wäre dies für Sprecher von Sprachen mit komplexen Silben nicht notwendig, da die einfacheren Silbenstrukturen von Silbensprachen auch in deren Muttersprache vorkommen (vgl. den Os, 1985; 1988). Sofern in diesen eine höhere Silbengeschwindigkeit vorliegt, was aufgrund der einfacheren Silben in den meisten Fällen zu erwarten ist, wird auch ein höheres Tempo wahrgenommen. Nach Pfitzingers Definition der wSG ist klar, dass im Sprachvergleich voneinander abweichende Silben- oder Lautgeschwindigkeiten die wSG erklären können, auch, wenn die andere der beiden Geschwindigkeiten in den miteinander verglichenen Sprachen gleich ist. Der Unterschied zwischen objektiv gemessener SG und subjektiver wSG kann dabei je nach Beschaffenheit der Stimuli und der Muttersprache der Hörer unterschiedlich stark ausfallen. Pfitzinger und Tamashima stellen fest, dass Deutsche das Sprechtempo des Japanischen stärker überschätzen (9,1%) als Japaner das des Deutschen (7,5%).

¹⁷⁵ Die Autoren untersuchen die Qualität illusorischer Vokale bei Sprechern des BP und Japanischen und zeigen, dass wahrgenommene Sprossvokale vor der lexikalischen Verarbeitung der Stimuli eingefügt werden, vgl. Dupoux et al. (1999, 2011).

¹⁷⁶ Das Ergebnis entspricht dem in den Os (1988).

Anschließend werden Faktoren beschrieben, denen ein Einfluss auf die wSG, meist in der Muttersprache, nachgewiesen wurde.

5.4 Faktoren, die das wahrgenommene Sprechtempo beeinflussen

5.4.1 Grundfrequenz

Sowohl der relative Frequenzbereich als auch die Dynamik des f_0 -Verlaufs beeinflussen die wSG: Eine hohe Grundfrequenz geht mit schnellem wahrgenommenem Tempo, eine niedrige Grundfrequenz mit langsamem wahrgenommenem Tempo einher (Hoequist/Kohler, 1986a); monotone Äußerungen erwecken den Eindruck höherer Geschwindigkeit (Hoequist/Kohler, 1986a; den Os 1985; 1988: 86), was mit der objektiv gemessenen AG zusammenhängt, die in Äußerungen mit dynamischem f_0 -Verlauf niedriger ist. Gleichzeitig gilt auch generell, dass monotone f_0 -Verläufe die wSG weniger beeinflussen als dynamische, d.h. dass z.B. ein dynamischer niedriger f_0 -Verlauf den Eindruck niedriger SG verstärkt.¹⁷⁷ Kohler (1986b: 168) ergänzt diese Erkenntnisse: Ein finaler Abfall der f_0 bewirkt eine Verringerung des wahrgenommenen Tempos, umgekehrt resultiert ein Anstieg der f_0 zum Satzende hin in einer wahrgenommenen Beschleunigung.¹⁷⁸ Über ihre Wahrnehmungsexperimente berichtend erwähnen Rietveld und Gussenhoven (1987: 281) eine Untersuchung von Sundberg (1979), nach der ein Abfall der f_0 schneller produziert werden kann als ein Anstieg, merken jedoch an, dass keine Beweise dafür vorliegen. In ihrer Untersuchung zeigen die Autoren, dass Äußerungen, in denen ein Ton infolge der „Tone Linking Rule“ ($T \rightarrow \emptyset / _ \dot{T}$) vor einem Hauptton getilgt wird, eine höhere wahrgenommene SG auslösen. Dabei entfällt z.B. nach einem hohen Ton der Abfall und anschließende erneute Anstieg der f_0 vor einem weiteren hohen Ton, was in diesem Fall in einem f_0 -Plateau resultiert (Bsp. s. Rietveld/Gussenhoven 1987: 275). Eine mögliche Erklärung hierfür finden Rietveld und Gussenhoven in der fehlenden finalen Dehnung innerhalb der ersten tonalen Einheit, die infolge des Ausfalls der Grenze mit der zweiten tonalen Einheit zusammenfällt, also in der tatsächlichen Dauer und Geschwindigkeit.

In Anbetracht von Pfitzingers Ergebnissen muss davon ausgegangen werden, dass die lokale und die globale wSG auf unterschiedliche Weise von Tonhöhenveränderungen

¹⁷⁷ Vgl. jedoch Gósy (1992: 171), die für niedriges und mittleres Tempo eine schwache Tendenz im Ungarischen feststellt, monotone Sätze als langsamer zu bewerten.

¹⁷⁸ Kohler (1986b: 171) rechnet, im Gegensatz zu Pfitzinger, der f_0 eine besondere Gewichtung sowohl für die Perzeption als auch bei der Produktion des Tempos zu und stellt das „Aus zählen“ von Lauten oder Silben seitens Hörer zur Einschätzung des Tempos als unnötig dar.

beeinflusst werden. Darüber hinaus kann nicht mit Sicherheit gesagt werden, dass der Tonhöhenverlauf für Hörer unterschiedlicher Muttersprachen gleichermaßen einen Einfluss auf die wahrgenommene SG hat (relevant ist hierbei möglicherweise das Vorhandensein phonologischer Töne).

5.4.2 Pausen

Die Darstellungen in 2.4 ergänzend sollen hier Untersuchungen zur Flüssigkeit der Rede erwähnt werden (Shriberg, 1999; Lövgren/van Doorn, 2005), die zeigen, dass die wahrgenommene Flüssigkeit mit steigender Pausendauer abnimmt.¹⁷⁹ In Untersuchungen wie der von Lövgren/van Doorn (2005) werden jedoch manipulierte Nachrichtenaufzeichnungen untersucht, so dass ein künstlich erzeugtes Missverhältnis zwischen stillen Pausen und gesprochener Sprache als Stimuli präsentiert wird. Dabei kann nicht davon ausgegangen werden, dass natürlich erzeugte längere Pausen, die mit entsprechenden finalen Längungen und möglicherweise mit einem anderen, natürlicheren f₀-Verlauf produziert werden, den gleichen Effekt erzeugen.

In Bezug auf den Einfluss von Pausen auf die wSG stellt Koreman (2006a) fest, dass Stottern und gefüllte Pausen den Eindruck einer niedrigen SG erwecken können, wohingegen Dehnungen und Versprecher mit einer höheren wSG einhergehen. Anhand der Pausenhäufigkeit unterscheiden sich schnelle und langsame Leser eines Textes nicht, allerdings produzieren als schnell empfundene Sprecher beim Nacherzählen *mehr* Pausen als Sprecher, die subjektiv als langsam bezeichnet werden (vgl. 4.5.7, wo festgestellt wurde, dass häufigere Pausen einen schwachen Zusammenhang mit einem niedrigeren Gesamtpausenanteil aufweisen). Die Wahrnehmung von Pausen hängt wiederum von deren Länge und der globalen AG ab (s. 2.3). Wenn man davon ausgeht, dass eine mit vielen Pausen versehene Äußerung in mehr Einheiten (Sinneinheiten, sofern es sich um syntaktische Pausen handelt) untergliedert ist als eine mit weniger Pausen versehene, und diese stärker unterteilte Äußerung eben als schneller empfunden wird, so ergibt sich ein konträres Bild zu dem, was aus Untersuchungen zum Tonhöhenverlauf hervorgeht. Kohler (1986b: 173 f.) spricht der Pausendauer einen Einfluss auf die wSG deshalb ab, weil Pausen mit finaler Längung einhergehen und darüber hinaus die Tonhöhe über längere Pausen hinweg größere Sprünge machen kann und diese beiden Eigenschaften, finale Längung und f₀-Verlauf alleine, die wahrgenommene SG erklären können.

¹⁷⁹ Zwar darf flüssige Rede nicht mit hohem Tempo gleichgesetzt werden, allerdings kann ein Zusammenhang von nicht flüssiger Rede mit niedrigem wahrgenommenem Sprechtempo vermutet werden.

5.4.3 Artikulatorische Genauigkeit

Neben tatsächlich realisierten Lauten können zur Bestimmung der AG phonologisch zugrunde liegende Elemente gezählt werden. In einigen Arbeiten (z.B. Koreman, 2003; 2006b) werden diese beiden Lautgeschwindigkeiten miteinander verglichen.¹⁸⁰ Dieser Vergleich beruht auf der Annahme, dass durch den Sprecher phonetisch reduzierte Formen vom Hörer in ihrer phonologischen „Idealform“ wahrgenommen, also ergänzt werden. Da von Reduktionen und Elisionen gekennzeichnete Äußerungen weniger Sprachmaterial umfassen, können diese in einem kürzeren Zeitraum als die vollständige Laut- bzw. Silbenkette artikuliert werden und eine höhere wSG auslösen (sofern Hörer die Elisionen tatsächlich wahrnehmen – dies kann jedoch selbst für Sprecher in Frage gestellt werden). Dass die zeitliche Komprimierung durch Nullrealisierung einiger Phoneme so stark ausgeprägt ist, dass der Hörer einen merkbaren Unterschied in der absolut beanspruchten SZ wahrnimmt, ist dagegen fragwürdig, v.a. unter Berücksichtigung der Wahrnehmungsschwelle von fünf Prozent bei Sprechtemposchwankungen (Quené, 2007). Es wird außerdem vorausgesetzt, dass ein direkter Vergleich mit der phonologischen Idealform der Äußerungen stattfindet. In Koreman (2003) vergleichen 20 Hörer zwei Aufnahmen in je 75 Paaren. Es liegen tatsächlich Bewertungen als „unterschiedlich schnell“ vor, wenn sich die Aufnahmen nur bezüglich der reellen Lautgeschwindigkeit oder aber nur in Hinblick auf die intendierte SG (kanonische Laute/s) unterscheiden. Werden phonetisch kaum reduzierte mit phonetisch stark reduzierten Äußerungen gleicher AGs verglichen, so wird in ungefähr der Hälfte der Fälle die Aufnahme mit der höheren intendierten Geschwindigkeit als schneller empfunden. Dies belegt, dass sowohl die rein auditive Wahrnehmung sprachlicher Stimuli als auch die Rekonstruktion zugrunde liegender phonologischer Formen die wSG beeinflussen (der Inhalt muss dabei nicht notwendigerweise verstanden werden). Bei gleichen intendierten Geschwindigkeiten wird tendenziell die Aufnahme mit der höheren realisierten Geschwindigkeit als schneller bezeichnet. Weichen beide Lautgeschwindigkeiten voneinander ab, wird in der Mehrheit der Fälle ein Unterschied wahrgenommen. Allerdings werden auch Stimuli, die hinsichtlich der tatsächlichen und der intendierten AG gleich sind in der Hälfte der Fälle als ungleich schnell bezeichnet. Koreman (2006b) bestätigt die Ergebnisse.¹⁸¹

¹⁸⁰ Neben den Begriffen „tatsächlich realisiert“ bzw. „effektiv“ und „kanonisch“ werden auch „phonetisch“ und „phonologisch“, „tatsächlich“ und „fiktiv“, „brutto“ und „netto“, von den Os (1988) auch „Silbenzahl“ und „linguistische Silbenzahl“ oder von Koreman „realisiert“ und „intendiert“ verwendet.

¹⁸¹ Weiterhin sei hier wiederholt, dass die tatsächliche AG und die artikulatorische Genauigkeit prinzipiell als unabhängig voneinander gelten (vgl. 2.3.1). Krause/Braida (2002) zeigen, dass deutliche Aussprache bei

5.4.4 Silbenstrukturen

Hoequist/Kohler (1986: 20 f.) prognostizieren, dass eine komplexe Silbe schneller klingt als eine einfache Silbe, wenn diese gleiche Dauern aufweisen und hintereinander präsentiert werden, was in einem Wahrnehmungsexperiment als Tendenz nachgewiesen werden kann und auf die Lautgeschwindigkeit zurückzuführen ist. Ob dieses Ergebnis für die Wahrnehmung natürlicher gesprochener Sprache von Bedeutung ist, ist jedoch unsicher, da komplexere Silben nicht zwingend mit einer deutlich höheren Lautgeschwindigkeit produziert werden müssen.¹⁸² Für einen Sprachvergleich ist die Frage interessant, ob eine Abfolge von vielen überwiegend komplexen Silben anders wahrgenommen und in Bezug auf das Sprechtempo eingeschätzt wird als eine Abfolge überwiegend einfacher Silben (ohne Manipulation deren Dauer). Es ist davon auszugehen, dass, wie oben beschrieben, Muttersprachler von Sprachen, in denen CV-Abfolgen die Regel sind, illusorische Vokale ergänzen, damit die Anzahl der wahrgenommenen Laute sowie Silben/s erhöhen und die wahrgenommene SG bei diesen Hörern höher liegt als bei Sprechern von Sprachen, in denen viele komplexe Silben die Regel sind. Diese wiederum werden von der höheren Anzahl von Silben(kernen)/s einer Sprache mit ausgeprägter CV-Tendenz so beeinflusst, dass sie die Sprache als schneller empfinden. Aufgrund unterschiedlicher Korrelationen zwischen Laut- und Silbengeschwindigkeit, die auf verschiedene Silbenkomplexitäten zurückzuführen sind (s. 4.5.7.2), ist davon auszugehen, dass diese beiden Aspekte des Tempos mit verschiedenen Gewichtungen die wSG beeinflussen.

5.4.5 Erwartungshaltung (Stereotype)

Die in Kapitel 2 erwähnten Kommentare von Essens, Flachskampfs und Anderer verdeutlichen, dass Stereotype bezüglich der SG anderer (Sprach-)Gemeinschaften bestehen. Da kein lexikalischer Zugang möglich ist, wenn unvertraute Sprachen gehört werden, und segmentale Informationen von untrainierten Hörern nur schwer fassbar sind, können suprasegmentale Eigenschaften umso stärker in den Vordergrund treten. Dabei entstehen aufgrund des kategorisierenden Denkens und des für den Menschen unabdingbaren Urteilens Stereotype und Vorurteile,¹⁸³ wenn z.B. situative, pragmatische und weitere

verschiedenen SGs die Verständlichkeit unabhängig von der objektiv gemessenen SG erhöhen kann. Sofern auch in der Muttersprache geringe Verständlichkeit mit tendenziell höherer wSG einhergeht (s.o.), kann ein Zusammenhang von artikulatorischer Genauigkeit und wSG erklärt werden.

¹⁸² In welchem Verhältnis steigende Silbendauer bei zunehmender Anzahl an Lauten in der Silbe und Kompression innerhalb der Silbe zueinander stehen, kann nicht pauschalisiert werden.

¹⁸³ Vgl. Kalverkämper (2000) zu kulturellen Vorurteilen.

Gegebenheiten unberücksichtigt bleiben und kürzere fremdsprachliche Äußerungen mit den eigenen Sprechgewohnheiten verglichen werden. Diese sind deshalb von Bedeutung, weil sie eine Erwartungshaltung auslösen, wenn das Tempo von Fremdsprachen eingeschätzt wird. Diese Erwartungshaltung wiederum kann die wSG beeinflussen, v.a. dann, wenn im fremdsprachlichen Kontext Probleme in der Kommunikation auftauchen. Ein „zu hohes Sprechtempo“ dient außerdem als Erklärung, wenn defizitäre Kenntnisse in oder eine nicht ausreichende Vertrautheit mit der Fremdsprache als Erklärung vermieden werden (was den Tatsachen entsprechen *kann*). Den Os (1988) erwähnt Meinungsumfrageergebnisse, nach denen Niederländer Italienisch als schnelle Sprache, Germanische dagegen als langsamere Sprachen empfinden (s. auch Kap. 7 zu Klischees, die die nonverbale Kommunikation betreffen). In 6.4 werden eigene Umfrageergebnisse vorgestellt.

5.4.6 Weitere Faktoren

Darauf, dass das Bewusstsein für die Lautgeschwindigkeit und einer Diskrepanz zwischen kanonischer und tatsächlicher Lautzahl vom sprachlichen Hintergrund der Hörer abhängig sein kann, weist das Ergebnis einer Untersuchung von Read et al. (1986) hin. Die Autoren zeigen, dass Muttersprachler des Chinesischen, die die alphabetische Schriftform ihrer Sprache nicht beherrschen, nicht in der Lage sind, Anweisungen nachzukommen, einzelne Laute in Silben zu ersetzen. Offensichtlich liegen einzelne Laute unter der sprachlichen Bewusstseinsgrenze von Sprechern, die keine alphabetische Schrift gelernt haben. Demnach ist nicht auszuschließen, dass für diese Hörer die Lautgeschwindigkeit eine geringere Bedeutung für die wSG hat.

Zwölf von sechs Phonetikstudenten subjektiv als schnell oder langsam bezeichnete Sprecher zeigen keine Unterschiede in ihrer Beurteilung verschiedener Aufnahmen (Koreman, 2006b), so dass nicht davon ausgegangen werden kann, die eigenen Sprechgewohnheiten hätten einen Einfluss auf die wSG.

In Untersuchungen von Feldstein et al. (1993) kann ein derartiger, anfänglich beobachteter Effekt nicht reproduziert werden. Dafür wird in einem Experiment mit 51 Teilnehmern, die die SG in kürzeren Aufnahmen auf einer Skala beurteilen, festgestellt, dass das Geschlecht einen Einfluss auf die wSG hat. Bei vergleichbaren objektiven SGs schätzen Frauen die Geschwindigkeiten der Aufnahmen höher ein als Männer. Gleichzeitig werden von allen Hörern weibliche Sprecher als langsamer empfunden. Allerdings werden in der Untersuchung pro Sprechergeschlecht nur drei Aufnahmen verwendet und keine genaueren

Angaben zur Beschaffenheit dieser gemacht (z.B. bezüglich f_0 , Intensität, etc.). Damit sollte das Geschlecht sowohl der Hörer als auch der Sprecher als *möglicher* Faktor der wSG festgehalten werden.

Für nicht verständliches Material können Feldstein/Bond (1981) und Bond/Feldstein (1982) Effekte der Tonhöhe als auch der Intensität der Stimuli auf die wSG nachweisen. In der jüngeren Untersuchung wird in neun Stimuluspaaren (rückwärts abgespielte von einer Frau gesprochene Beschreibungen eines Bildes) jeweils eine unmanipulierte Version einer gleich langen manipulierten Version gegenübergestellt, in der die Grundfrequenz ($\pm 12,9\%$) oder die Intensität (± 5 dB) und zudem die Geschwindigkeit verändert wurden. Die Paare werden drei Gruppen von Versuchsteilnehmern in drei verschiedenen SGs präsentiert (100, 150 und 200 Wörter/min). Auf vier sieben-Punkte-Skalen können die insgesamt 61 Teilnehmer angeben, ob und wie sehr sie die zweite, manipulierte Aufnahme in Bezug auf Tonhöhe, Lautstärke, Geschwindigkeit und Länge als abweichend einschätzen. In allen drei Gruppen, d.h. für alle drei SGs, kann neben einem signifikanten direkten linearen Zusammenhang zwischen der Tonhöhe und der wSG, der 28% der Varianz der SG-Urteile erklärt, ein entsprechender Effekt für die Intensität nachgewiesen werden, der allerdings nur 1,6% der Varianz erklärt.

5.5 Empirische Untersuchung zum wahrgenommenen Sprechtempo

In einem Wahrnehmungsexperiment wurde untersucht, ob bei der Einschätzung sprachlicher Stimuli systematische Unterschiede zwischen Muttersprachlern des Chinesischen und des Deutschen festzustellen sind. Dazu wurde in einer vorläufigen Untersuchung getestet, ob Muttersprachler des Chinesischen, die über keine Fremdsprachenkenntnisse verfügen, in ihrer Beurteilung des Tempos romanischer und germanischer Sprachen übereinstimmen.

5.5.1 Vorläufige Untersuchung

Acht nicht manipulierte, zufällig gewählte Ausschnitte aus den ausgewerteten Nachrichtensendungen wurden als Stimuli präsentiert. Dabei handelt es sich um Aufnahmen je eines Sprechers pro Geschlecht der Sprachen BP, Deutsch, Englisch und Spanisch. Die Aufnahmen umfassen eine Länge von etwa einer Minute, wurden aber von den Versuchsteilnehmern nur so lange angehört, bis sie ein Urteil bezüglich der SG fällen konnten. Das Material wurde über einen CD-Player mit externen, seitlich platzierten Boxen präsentiert, an dem die Teilnehmer die Lautstärke selbst regeln konnten. Das Urteil

wurde in Form einer Markierung in einer Tabelle eingetragen, in der nur die drei Kategorien „langsam“, „mittel“ und „schnell“ zur Auswahl standen. Wie erwartet wurde die Kategorie „langsam“ nicht verwendet. Die Bezeichnungen „mittel“ und „schnell“ werden deshalb als „relativ langsam“ und „relativ schnell“ gewertet. Obwohl nur zehn chinesische Teilnehmer unterschiedlicher Altersklassen und Berufe zu dem informellen Test eingeladen wurden und die Ergebnisse damit nicht generalisiert werden dürfen, kann eine eindeutige Tendenz ausgemacht werden: In mehr als der Hälfte der Aufnahmen (fünf von acht) stimmen mindestens acht Teilnehmer in ihrer Beurteilung überein. In Tabelle 5.1 werden die Aufnahmen ihrer chronologischen Abfolge bei der Präsentation entsprechend angeordnet.

Tabelle 5.1: Beurteilung der SG der Sprachen BP, DE, EN und ES von zehn Muttersprachlern des Chinesischen; ab hier bezeichnen auf ‘w’ endende Abkürzungen von Frauen gesprochene, auf ‘m’ endende von Männern gesprochene Aufnahmen

Aufnahme (Sprecher)	Beurteilung			
	Langsam	Mittel	Schnell	k.A.
DEw	0	4	6	
BPw	0	2	6	2
ESm	0	3	7	
ENw	0	10	0	
DEm	0	9	1	
ENm	0	9	0	1
BPm	0	0	10	
ESw	0	1	8	1

Da keiner der Teilnehmer über Fremdsprachenkenntnisse verfügt und keine der Sprachen erkannte, kann eine auf Stereotype zurückgehende Erwartungshaltung in diesem Fall ausgeschlossen werden. Die beiden romanischen Sprachen werden eindeutig als schneller empfunden. Um zu untersuchen, ob dieser Eindruck auf die häufigeren CV-Silben und die damit schneller wechselnde Schallfülle (höhere Anzahl von Silbenkernen/s) zurückzuführen ist, wurde ein Wahrnehmungsexperiment mit kürzeren Ausschnitten durchgeführt.

5.5.2 Wahrnehmungsexperiment zur Bedeutung der Silbenstruktur

5.5.2.1 Methode

Hierfür wurden Studenten eingeladen, die vergleichbare Voraussetzungen erfüllen. Die zu beurteilenden Stimuli stellen kurze Ausschnitte aus den untersuchten Nachrichtenaufnahmen dar. Durch die Verwendung mehrerer Sprachen, mit denen die Teilnehmer nicht vertraut sind (s.u.), sollten wieder Effekte, die auf Vorurteilen basieren, weitgehend vermieden werden. Das Experiment wurde in Sprachlaboren an PCs mit vergleichbarer Ausstattung durchgeführt. Auf einer grafischen Oberfläche wurden die Symbole der Audiodateien der einzelnen Aufnahmen vertikal übereinander in der Mitte der Fläche angeordnet. (s. Abb. 10.1 im Anhang). Die Teilnehmer wurden aufgefordert, jede einzelne Aufnahme mehrmals über Kopfhörer anzuhören und sie entsprechend ihrer Geschwindigkeit an den linken oder rechten Bildschirmrand zu verschieben und die Anordnung am Ende des Experimentes durch erneutes Abhören aller Aufnahmen zu überprüfen. Am unteren Bildschirmrand war ein doppelseitiger Pfeil zu sehen, der links auf den Begriff „langsam“, rechts auf den Begriff „schnell“ (in der Muttersprache der Versuchsteilnehmer) verwies. Die Teilnehmer wurden schriftlich und mündlich angewiesen, die Aufnahmen zueinander in Bezug zu setzen und die Bildschirmbreite auszunutzen. Nach Abschluss der Beurteilung wurde ein Gitternetz über die grafische Oberfläche mit der Anordnung jeder einzelnen Versuchsperson projiziert, anhand derer Werte für die einzelnen Aufnahmen ermittelt wurden. An der ursprünglichen Position der Dateien wurde dabei der Wert 0 platziert, zur rechten Seite positive und zur linken Seite negative Werte. Die Skala reicht von -6 bis +6 und die Werte wurden in 0,5-Schritten abgelesen. Auf diese Weise sollte vermieden werden, dass Teilnehmer extreme Urteile meiden und sich auf eine zuvor festgelegte Skalierung beschränken.

5.5.2.2 Teilnehmer

Es handelt sich um 75 chinesische und 76 deutsche Studenten in ihrem Heimatland, die ein nicht-sprachliches Fach (Wirtschaftswissenschaften bzw. ein technisches oder naturwissenschaftliches Fach) im zweiten bis vierten Semester studierten. Der Anteil weiblicher und männlicher Studenten ist nicht ganz ausgewogen: Die chinesische Gruppe bestand aus 37 Studentinnen und 38 Studenten, die deutsche aus 44 Studentinnen und 32 Studenten. Die deutschen Studenten verfügen teilweise über Fremdsprachenkenntnisse auf fortgeschrittener Stufe in Englisch und Grundkenntnisse einer weiteren Fremdsprache,

wohingegen die chinesischen Studenten neben Kenntnissen der englischen Sprache nur in wenigen Fällen über Grundkenntnisse einer meist ostasiatischen Fremdsprache verfügen.

5.5.2.3 Stimuli

Die untersuchten Sprachen BP, Chinesisch, Deutsch, Englisch, EP und Spanisch wurden jeweils in zwei Aufnahmen berücksichtigt, von denen eine eine Abfolge von einfachen Silben, die andere eine Abfolge von komplexen Silben darstellt. Alle Aufnahmen wurden auf 175 Hz monotonisiert um mögliche Effekte des f_0 -Verlaufs auszuschließen, die maximale Intensität wurde auf 80 dB festgelegt. Die tatsächlichen Lautgeschwindigkeiten der Stimuli liegen zwischen 10,59 und 15,51 Lauten/s. Deutlichere Unterschiede bestehen bezüglich der Silbengeschwindigkeit, die zwischen 2,8 und 6,8 Silben/s liegt; die Silbenkomplexität liegt zwischen 2,0 und 4,8 Lauten/Silbe. Die Länge der Aufnahmen fällt mit 0,8 – 1,6 s unterschiedlich aus. Siehe Abbildung 5.1 zu Laut- und Silbengeschwindigkeiten der Stimuli.

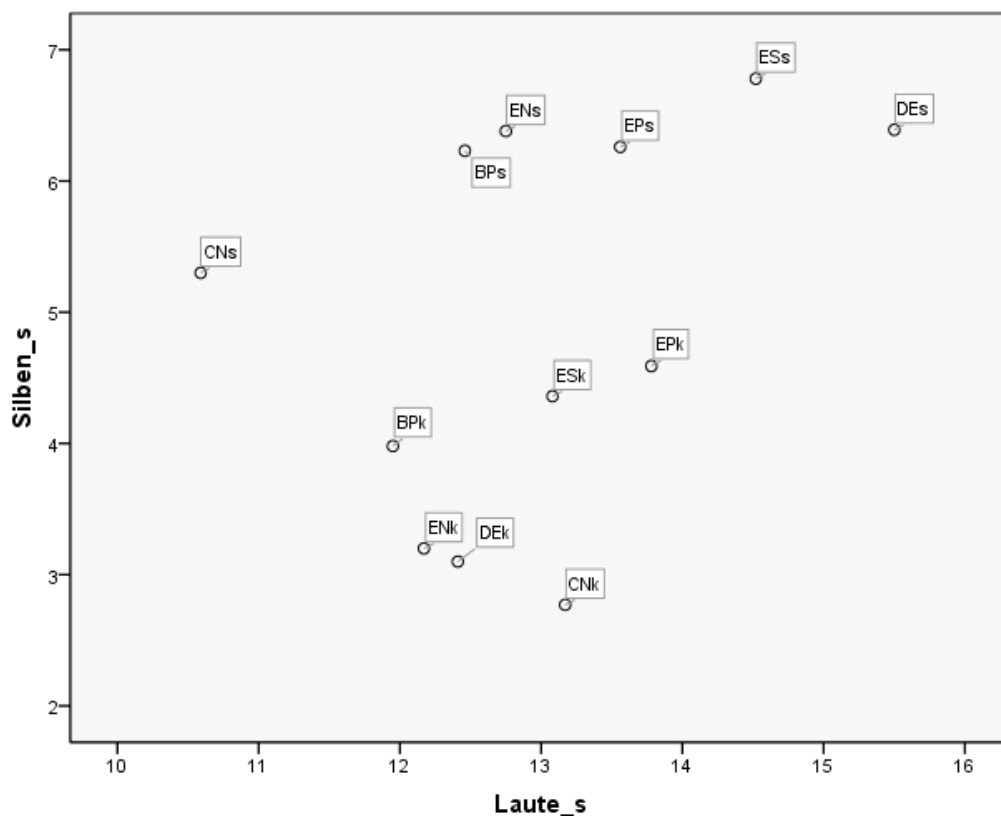


Abbildung 5.1: Laut- und Silbengeschwindigkeiten der Stimuli des Wahrnehmungsexperiments (auf „k“ endende Stimuli stellen Abfolgen komplexer Silben dar, auf „s“ endende Stimuli stellen Abfolgen einfacher Silben dar)

5.5.2.4 Ergebnisse

Es können signifikante Unterschiede bezüglich der Bewertung der Stimuli zwischen chinesischen und deutschen Muttersprachlern festgestellt werden. Innerhalb der einzelnen Sprachen der Aufnahmen wird die Abfolge komplexer Silben im Vergleich mit der einfacher Silben von den deutschen Muttersprachlern öfter als langsamer bewertet als von den chinesischen Muttersprachlern, d.h. unter den deutschen Versuchsteilnehmern wird die Abfolge einfacher Silben im Vergleich mit der Abfolge komplexer Silben derselben Sprache häufiger als schneller empfunden. Die durchschnittliche Skalierung (=Mittelwerte der Beurteilungen) zeigt, dass zwischen den Abfolgen einfacher Silben aller Sprachen und den Abfolgen komplexer Silben aller Sprachen für deutsche Hörer ein größerer Unterschied besteht ($0,62 > -1,14$) als für chinesische Hörer ($0,3 > -0,81$). Sowohl bei der durchschnittlichen Bewertung aller Abfolgen komplexer Silben ($T_{(149)}=2,22$; $p=0,028$), als auch bei der einfacher Silben ($T_{(149)}=-2,064$; $p=0,041$) sind signifikante Unterschiede zwischen den Teilnehmergruppen festzustellen. In Hinblick auf die Anzahl der Paare (komplexe und einfache Silben einer Sprache), in denen komplexe Silbenketten als langsamer empfunden wurden, liegt ein hochsignifikanter Unterschied vor ($T_{(149)}=-4,117$; $p<0,01$). Die Einschätzung der Geschwindigkeit der einzelnen Aufnahmen ist in acht von zwölf Fällen signifikant unterschiedlich.¹⁸⁴

Die Korrelationsmatrix in Tabelle 5.2 zeigt, dass die Wahrnehmung deutscher Muttersprachler stark negativ mit der Silbenkomplexität korreliert ($r=-0,835^{**}$), d.h. je mehr Laute die Silben der Aufnahmen enthalten, desto langsamer wird das Tempo geschätzt. Diese Korrelation ist bei den chinesischen Muttersprachlern schwächer und nicht signifikant ($r=-0,542$). Die Silbengeschwindigkeit korreliert unter deutschen Teilnehmern stärker mit der wSG ($r=0,410$) als die Lautgeschwindigkeit ($r=0,231$); unter chinesischen Teilnehmern korreliert umgekehrt die Lautgeschwindigkeit ($r=0,482$) stärker als die Silbengeschwindigkeit ($r=0,394$) mit der wSG.

Tabelle 5.2: Korrelationsmatrix des Wahrnehmungsexperiments zwischen WG (= gemittelte Werte auf Wahrnehmungsskala für die einzelnen Aufnahmen, von deutschen (de) und chinesischen (cn) Muttersprachlern) und phonetischer Beschaffenheit der Stimuli

		Silbenkomplexität	Laute/s	Silben/s	Dauer
WGde	r	-,835 ^{**}	0,231	0,410	-0,709 ^{**}
	p	0,001	0,470	0,186	0,010
WGcn	r	-0,542	0,482	0,394	-0,472
	p	0,069	0,112	0,206	0,121

¹⁸⁴ CNS^{**}, CNk^{**}, DES^{*}, DEk^{**}, EPS^{**}, EPk^{*}, ESS^{**}, ESK^{**}.

Zusammenfassend muss festgehalten werden, dass die wSG der Muttersprachler des Chinesischen von anderen Eigenschaften beeinflusst wird als die der Muttersprachler des Deutschen. Silbenkomplexität und Silbengeschwindigkeit sind für die beiden Gruppen von unterschiedlicher Bedeutung. Für Muttersprachler des Chinesischen scheint die Erklärung plausibel, dass in Konsonantenhäufungen perzipierte Silbenkerne vorliegen, die die wahrgenommene Silbengeschwindigkeit gegenüber der objektiv gemessenen erhöhen und damit einen schwächeren Zusammenhang zwischen objektiver Silbengeschwindigkeit und wSG verursachen als zwischen Lautgeschwindigkeit und wSG. Die Silbenkomplexität weist für beide Gruppen den stärksten Zusammenhang mit der wSG auf. Für chinesische Teilnehmer ist dieser negative Zusammenhang („je mehr Laute pro Silbe, desto langsamer klingt eine Aufnahme“) jedoch weitaus weniger stark ausgeprägt, da eine durch viele Konsonanten erhöhte Silbenkomplexität die wSG wiederum erhöht. Von diesem Effekt können durchaus nur manche Konsonantengruppen betroffen sein: Unter Berücksichtigung der Ergebnisse in 4.5.7.5 gilt als wahrscheinlich, dass monosegmental produzierte „Lautgruppen“ wie [ts^h] auch monosegmental wahrgenommen und nicht durch perzipierte Vokalkerne zu wahrgenommenen Silben ergänzt werden. Es kann daher auch keine positive Korrelation zwischen Silbenkomplexität und wSG festgestellt werden. Die Einflüsse auf die wSG chinesischer Teilnehmer lassen sich nur als Tendenzen erkennen und weitere Faktoren dürfen nicht ausgeschlossen werden. So ist es nicht auszuschließen, dass auch die absolute Dauer der Aufnahmen eine Wirkung auf die wSG hat, worauf eine starke negative signifikante Korrelation bei deutschen Hörern hinweist. Hinweise auf einen kausalen Zusammenhang sind jedoch nicht bekannt.

Ein lineares Regressionsmodell (Einschluss signifikanter Variablen) zur Erklärung der Faktoren der wSG deutscher Versuchsteilnehmer liefert nur die Silbengeschwindigkeit als signifikanten Faktor (zur Auswahl gestellte Faktoren: Silbenkomplexität, Laut- und Silbengeschwindigkeit und Dauer):

$$wSG_{de} = 0,718 \cdot \text{Silbengeschwindigkeit}; R^2=0,725; p<0,001.$$

Unter Einbezug weiterer Faktoren verbessert sich die Güte nur geringfügig. Das beste Modell zur Erklärung der wSG chinesischer Versuchsteilnehmer schließt die Faktoren Laut- und Silbengeschwindigkeit und Silbenkomplexität ein:

$$wSG_{cn} = -4,115 \cdot \text{Silbenkomplexität} + -2,01 \cdot \text{Silbengeschwindigkeit} + 1,524 \cdot \text{Lautgeschwindigkeit}; R^2=0,564; p=0,072.$$

Eine nur geringfügig niedrigere Güte von $R^2=0,499$ wird unter Einschluss der Faktoren Lautgeschwindigkeit und Silbenkomplexität berechnet. Damit ist davon auszugehen, dass

neben Silben- und Lautgeschwindigkeit sowie Silbenkomplexität noch andere Faktoren existieren, die in der Untersuchung außer Acht gelassen wurden.

5.6 Zusammenfassung

Untersuchungen zeigen, dass die wSG von der Pausenstruktur und der artikulatorischen Genauigkeit beeinflusst wird. Weiterhin wird die Grundfrequenz als Faktor angenommen, auch wenn ihr Einfluss auf die lokale wSG zumindest im Deutschen nur gering ist. Das Geschlecht sowohl der Hörer als auch der Sprecher ist als möglicher weiterer Faktor festzuhalten und für die Intensität ist ebenfalls ein geringer Einfluss nachweisbar. Auch die Variable „Sprache“ (einer bezüglich des Tempos beurteilten Äußerung) kann als verdeckter Faktor der wSG bezeichnet werden, da eine durch Stereotype bedingte Erwartungshaltung durchaus Einflüsse auf die wSG aufweisen kann (vgl. folgendes Kapitel). Silben- und Lautgeschwindigkeit sind beide für die wSG grundlegend, wobei die Bedeutung der Lautgeschwindigkeit möglicherweise von der Alphabetisierung der Hörer abhängt. Im Allgemeinen muss unterschieden werden, ob die Muttersprache, eine bekannte oder völlig unbekannte Fremdsprache beurteilt wird. Dabei können unterschiedliche Strategien zur Beurteilung des Tempos zugrunde liegen, die zumindest im letzten Fall der Kompatibilität des Stimulus mit phonotaktischen Verhältnissen in der Muttersprache der Hörer unterliegen. Hier ist an Silbenstrukturen als Faktor zu denken.

Das Wahrnehmungsexperiment hat gezeigt, dass die Silbenkomplexität für deutschsprachige Hörer viel stärker von Bedeutung ist als für chinesischesprachige, was mit den unterschiedlichen phonotaktischen Systemen der beiden Sprachen zusammenhängt. Perzeptive Silbennuklei erhöhen die wahrgenommene Silbengeschwindigkeit, wenn mit dem phonologischen System der Muttersprache inkompatible Konsonantenhäufungen in den Stimuli enthalten sind. Die Muttersprache der Hörer ist also ein entscheidender Faktor der wSG.

Nachdem das Sprechtempo in den Bereichen Produktion und Perzeption von einem phonetisch-phonologisch Standpunkt aus betrachtet wurde, soll im anschließenden Teil der Arbeit behandelt werden, inwiefern gesellschaftlich festgelegte Normen das bevorzugte Sprechtempo beeinflussen können. Dabei werden soziale Konventionen untersucht, die den Gebrauch verbaler wie nonverbaler sprachlicher Mittel beeinflussen.

Zweiter Teil

6 Kulturelle Aspekte des Sprechtempos

6.1 Einleitung

Die Schnelligkeit gesprochener Sprache wurde im ersten Teil dieser Arbeit linguistisch betrachtet. Daneben kann eine weitere Perspektive eingenommen werden, die auf einer Berücksichtigung zwischenmenschlicher Aspekte beruht. Es kann untersucht werden, mit welchen unterschiedlichen Konnotationen verschiedene Sprechtempi in gesellschaftlichen Gruppen einhergehen.¹⁸⁵ Zu diesem Zweck wird in diesem Kapitel nach einer einleitenden Begriffsklärung (6.2) eine Auswahl an Studien vorgestellt, in denen Sprecher auf unterschiedliche (Charakter-)Eigenschaften hin beurteilt werden, wobei das Sprechtempo dieser Sprecher als eine Variable untersucht wird (6.3). Darüber hinaus ist eine Spekulation über Zusammenhänge bevorzugter Sprechtempi mit weiteren Variablen, die den Kommunikationsstil beschreiben, von Interesse. Deswegen werden anschließend die Ergebnisse des vierten Kapitels mit kulturwissenschaftlich orientierten Untersuchungen verglichen, in denen anhand kultureller Dimensionen Beschreibungsmerkmale zur Abgrenzung von Kulturen herausgearbeitet werden (6.4). Abschließend werden eine Umfrage zu kulturellen Stereotypen (6.5) und ein Wahrnehmungsexperiment zur Bewertung von Sachlichkeit und Vertrauenswürdigkeit in Nachrichtensendungen (6.6) beschrieben. Die Zusammenfassung leitet zur Untersuchung nonverbaler Kommunikationsmittel im Sprachvergleich der beiden folgenden Kapitel über.

6.2 Definition des Kulturbegriffs

Zunächst muss geklärt werden, was unter dem häufig verwendeten und für diesen Teil der Arbeit grundlegenden Begriff „Kultur“ verstanden wird, für den keine allgemeingültige Definition vorliegt. Eine Definition, die der dieser Arbeit zugrunde liegenden Vorstellung einer relativ homogenen Gruppe von Menschen entspricht, die einen sprachlichen, d.h. verbalen und nonverbalen Zeichenkode teilt, ist die der „communities of practice“:

¹⁸⁵ Auch die Bedeutung von Schweigen ist hier ein relevanter Untersuchungsgegenstand, wie Stanienda (2002: 9 ff.) in ihrer Beschreibung der an unterschiedlichen Pausen- und *Turn*strukturen festgemachten Unterscheidung von schweigenden und sprechenden Kulturen feststellt; s. auch Sifianou (1997).

„A community of practice is an aggregate of people who come together around mutual engagement in some common endeavor. Ways of doing things, ways of talking, beliefs, values, power relations – in short practices – emerge in the course of their joint activity around that endeavor.“ (Eckert/Mc Connel-Ginet 1998: 490)¹⁸⁶

Matsumotos (2006: 220) Definition von „Kultur“, *„a shared system of socially transmitted behavior that describes, defines, and guides people’s ways of life, communicated from one generation to the next“*, kann als das Produkt konventionalisierter Normen von *communities of practice* angesehen werden. Auf den Bereich des Sprachgebrauchs eingegrenzt definiert Gumperz (1972[1968]: 114) den Begriff „Sprachgemeinschaft“ als „jedes menschliche Aggregat, das durch regelmäßige und häufige Interaktion mit Hilfe eines geteilten Vorrats an sprachlichen Zeichen charakterisiert ist und sich von ähnlichen Aggregaten durch signifikante Unterschiede im Sprachgebrauch abgrenzt“¹⁸⁷. Demnach können Land- und Stadtbevölkerung (vgl. 2.3.2) durchaus als verschiedene Sprachgemeinschaften bezeichnet werden, die sich im Gebrauch verbaler wie nonverbaler Kommunikationsmittel unterscheiden. Sprachgemeinschaften sind je nach Kontext auf diversen Ebenen voneinander zu unterscheiden, so kann die spanische der italienischen Sprachgemeinschaft, aber auch die ländliche Sprachgemeinschaft der Orkney-Inseln der städtischen Sprachgemeinschaft in Edinburgh (die sich wiederum unterteilen lässt) gegenüber gestellt werden. Problematisch ist dabei, dass Unterschiede oft so subtil sind, dass keine klaren signifikanten Abgrenzungen zwischen Gruppen erkennbar sind. Hilfreich scheint für die Abgrenzung von Sprachgemeinschaften deshalb die Empfindung der kommunizierenden Personen selbst als Zugehörige zur selben oder verschiedenen (Sprach-) Gemeinden.¹⁸⁸ Die Begriffe „Kultur“ und „kulturell“ werden in dieser Arbeit übernommen, weil sie in vielen Fachbereichen als etabliert gelten und im Kontext sprachlicher Unterschiede verwendet werden, die mit Unterschieden im kommunikativen Verhalten verschiedener gesellschaftlicher Gruppen oder Systeme einhergehen. Hier wird der Begriff in Anlehnung an obige Bedeutungen verwendet und damit impliziert, dass, auch wenn argumentiert werden kann, Brasilianer und Portugiesen sprechen die gleiche Sprache, sie sich doch in ihrem sprachlichen Verhalten unterscheiden, weil in den gesellschaftlichen Gruppen, die sie bilden, unterschiedliche Vorstellungen vorliegen, die in unterschiedlichen gesellschaftlichen sprachlichen Normen resultieren. Gleichzeitig können Chinesen, die nicht dieselbe Muttersprache sprechen, aufgrund geteilter gesellschaftlicher Normen eine

¹⁸⁶ Zitiert aus Raith (2004: 156), worauf auch die folgenden Überlegungen weitgehend beruhen.

¹⁸⁷ Zitiert aus Raith (2004: 151).

¹⁸⁸ Vgl. Mattheier (1980: 60), zitiert in Raith (2004: 149).

relativ einheitliche Gruppe¹⁸⁹ bilden. Der Begriff „Kultur“ wird außerdem verwendet um eine Abgrenzung gegenüber der Natur zu betonen, in der universale menschliche Verhaltensweisen gründen (vgl. 7.4).

6.3 Bisherige Untersuchungen

Zur Bedeutung der SG im gesellschaftlichen Kontext sind kaum umfassende kulturvergleichende Untersuchungen bekannt. Hall/Hall (1990) treffen eine überwiegend auf Überlegungen und persönlicher Erfahrung basierende Unterscheidung zwischen Kulturen, in denen schnell, und solchen, in denen langsam kommuniziert wird. Die Unterscheidung ist praktisch identisch mit der in *High-Context* (HC) und *Low-Context* (LC) – Kulturen.¹⁹⁰ In Ersteren, in denen der Kontext von Äußerungen als (mindestens) ebenso entscheidend für die Bedeutung der Äußerung gelten wie der Wortlaut selbst, werden Informationen schneller verbreitet, weil sie nicht explizit im Wortlaut, sondern implizit im Kontext enthalten sind, der jedem zugänglich ist, wohingegen in Letzteren der verbal kommunizierte Inhalt (eben der Wortlaut) von Äußerungen als eigentliche Information verstanden wird, die nur dann mitgeteilt wird, wenn sie explizit geäußert wird. Dieser Vorstellung kann entgegen gehalten werden, dass für Individuen aus LC-Kulturen Informationen, die dem Kontext entnommen werden und erst konstruiert werden müssen, als langsam empfunden werden, zumindest langsamer im Vergleich zu unmittelbar formulierten Informationen.¹⁹¹ Ob prinzipiell schnell oder langsam bzw. direkt oder indirekt kommuniziert wird, ist aber auch stark situationsabhängig und Pauschalisierungen treffen oft nicht zu.

In den USA liefern einige relevante Untersuchungen zur Persönlichkeitswahrnehmung interessante Ergebnisse, die weiter unten eingeschränkt werden. MacLachlan (1982) empfiehlt aufgrund von Erkenntnissen in dieser Forschungsrichtung für Werbesendungen beschleunigte Aufnahmen, da diese in einer Reihe von Untersuchungen positive Urteile bezüglich Vertrauenswürdigkeit und Wissen, Intelligenz und Kompetenz des Sprechers hervorrufen (s. auch Floyd, 1985: 54 f.). Miller et al. (1976) z.B. spielen 359 freiwilligen

¹⁸⁹ Kulturen sind selbstverständlich nicht als homogene Gebilde zu verstehen. Wie Matsumoto et al. (1996) deutlich veranschaulichen, ändern sich Werte in Kulturen, was zu Veränderungen ihrer Positionierung in den kulturellen Dimensionen (nach Hofstede, 1980) wie z.B. Individualismus-Kollektivismus führt.

¹⁹⁰ Nach Hall (1976).

¹⁹¹ So kann es mindestens für Personen aus LC-Kulturen, aber genauso für Personen aus HC-Kulturen ein längerer Prozess sein, eine Ablehnung (z.B. einer Einladung oder eines Angebotes) als tatsächliche Ablehnung (nicht etwa als Ausdruck von Höflichkeit) zu dekodieren.

Passanten eine Aufnahme vor, in der die negativen Seiten des Genusses von Kaffee dargelegt werden. Die natürlich produzierten Aufnahmen können ihrem Tempo entsprechend in drei Gruppen unterteilt werden. Nach Präsentation der Aufnahme geben die Teilnehmer den Grad ihrer Zustimmung (1-10) zu Thesen auf einem Fragebogen an. In den schnelleren Aufnahmen wird der Sprecher als sachkundiger und überzeugender empfunden. In Smith et al. (1975) nimmt die wahrgenommene Kompetenz bei künstlich manipulierten Aufnahmen mit steigender Geschwindigkeit zu, für wahrgenommenes Wohlwollen ergibt sich ein präferiertes mittleres Tempo (Inhalt: einfache Aussagen). Ebenso werden in Apple et al. (1979) für eine unmanipulierte Version die höchsten positiven Urteile gefunden, erhöhtes Tempo wirkt sich weniger nachteilhaft auf die Glaubwürdigkeit aus als verringertes Tempo. In einem weiteren Experiment finden die Autoren heraus, dass auch Überzeugungskraft, Flüssigkeit der Rede und Nachdruck bei unmanipuliertem Sprechtempo als am größten, bei verringertem Tempo als gering bewertet werden, wobei möglicherweise auch das Thema der Aufnahmen (Meinung zu Minderheitenquoten an Universitäten bzw. Vorstellung, wie man einen Geldgewinn verwenden würde) einen Einfluss auf die Bewertung hat. Street et al. (1983) arbeiten mit natürlich produziertem Sprachmaterial unterschiedlicher Geschwindigkeiten und können anhand vielzähliger Beurteilungen durch 96 Teilnehmer feststellen, dass wahrgenommene Kompetenz und soziale Attraktivität bei langsamem Tempo signifikant geringer ausfallen. Von Bedeutung ist dabei die Sprechgewohnheit der Hörer: Wenn das gewohnte Sprechtempo der Hörer unterschritten oder deutlich übertroffen wird, fallen Bewertungen negativ aus. Die Untersuchung gibt auch Hinweise darauf, dass niedriges Tempo mit Formalität assoziiert wird: In einer formellen Sprechsituation (Bewerbungsgespräch) werden für niedrige Geschwindigkeiten weniger negative Urteile der Art „Die Sprechweise ist inakzeptabel langsam“ abgegeben als in einer informellen Sprechsituation (Unterhaltung). Im vierten Kapitel wurde festgestellt, dass in englischen und deutschen Nachrichtensendungen vergleichsweise langsam gesprochen wird, weshalb die Vermutung nahe liegt, dass in diesen drei Gesellschaften (in den USA, Großbritannien und Deutschland) niedriges Tempo mit Formalität verbunden wird. Woodall/Burgoon (1983) sehen auch Zusammenhänge von natürlich produzierter erhöhter Geschwindigkeit und wahrgenommener Extrovertiertheit, jedoch keine signifikanten Effekte für wahrgenommene Kompetenz und keinen Zusammenhang mit Überzeugungskraft. Gleichzeitig wird in der Studie eine höhere SG häufiger als ablenkend bezeichnet, hat jedoch keinen Einfluss auf die Erinnerung der Teilnehmer an den Inhalt.

Die Allgemeingültigkeit der in zahlreichen Untersuchungen festgestellten tendenziell negativen Bewertungen in Bezug auf Überzeugungskraft, Wohlwollen, Kompetenz, Glaubwürdigkeit, Sachkenntnis oder soziale Attraktivität bei niedriger SG wird spätestens 1992 in Frage gestellt, als ein Vergleich der Beurteilung der und Reaktion auf die SG von US-amerikanischen und koreanischen Hörern von Lee und Boster unternommen wird. Die Autoren gehen davon aus, dass in der koreanischen Sprachgemeinschaft niedriges Sprechtempo deshalb Kompetenz vermittelt, weil es mit besonders sorgfältig überlegter Wortwahl assoziiert wird und diese zur in der koreanischen Kultur wichtigen Wahrung des (eigenen oder fremden) Gesichts beiträgt. Diesem Gedankengang liegt eine soziale Interpretation des Begriffs „Kompetenz“ zugrunde: Nicht fachspezifisches Sachwissen, sondern die gesellschaftlich anerkannte Geschicklichkeit, peinliche Situationen für alle Beteiligten zu vermeiden, wird hier unter „Kompetenz“ verstanden. Streng genommen darf nicht davon ausgegangen werden, dass in der zum Vergleich untersuchten US-amerikanischen Gruppe dieses Konzept von „Kompetenz“ geteilt wird. Lee und Boster sprechen von kulturellen Unterschieden zwischen Korea und den USA, die sie an der Unterscheidung zwischen kollektivistisch und individualistisch geprägten Gesellschaften festmachen: Da in kollektivistisch geprägten Gesellschaften ein größerer Wert auf die Betonung und Förderung der Harmonie in zwischenmenschlichen Beziehungen gelegt wird als in individualistisch geprägten, trägt eine niedrige SG, die sorgfältige Wortwahl signalisiert, zur Wahrnehmung von Kompetenz bei. In ihrer Untersuchung werden 120 koreanische und 120 US-amerikanische Studenten in je drei Gruppen eingeteilt, die eine Aufnahme in unterschiedlichen Geschwindigkeiten hören (langsam, normal, schnell). Die Aufnahme wird außerdem von einer Sprecherin und einem Sprecher gesprochen¹⁹² und stellt inhaltlich eine Nachricht dar, die das Publikum mit von statistischen Informationen untermauerten Argumenten vom Rauchen abhalten soll. Anschließend füllen die Teilnehmer Fragebögen zur Kompetenz, Vertrauenswürdigkeit und sozialen Attraktivität der Sprecher aus. SG und Geschlecht werden somit von den einzelnen Versuchsteilnehmern nicht miteinander verglichen. Für die wahrgenommene Kompetenz ist unter US-Amerikanern eine Interaktion von Geschlecht und SG zu erkennen: Der Sprecher männlichen Geschlechts wird mit hoher SG als kompetenter bewertet (signifikanter Unterschied), für die Sprecherin liegt kein Einfluss der SG auf die wahrgenommene Kompetenz vor. Für Vertrauenswürdigkeit liegen gleiche, für soziale Attraktivität ähnliche

¹⁹² Die Autoren spezifizieren nicht, welche Gruppe von Studenten die Aufnahme der Sprecherin und welche die des Sprechers hörten, es scheint aber, dass innerhalb jeder Gruppe beide Aufnahmen verschiedenen Teilnehmern präsentiert wurden.

Ergebnisse vor. Die Ergebnisse der koreanischen Teilnehmer zeigen, dass Kompetenz, Vertrauenswürdigkeit und soziale Attraktivität des Sprechers bei niedriger SG höher eingeschätzt werden (signifikanter Unterschied), die Sprecherin dagegen wird bei niedriger SG als weniger kompetent bewertet; auf die wahrgenommene Vertrauenswürdigkeit und soziale Attraktivität hat die SG hier keinen Einfluss. Die Autoren ziehen die Schlussfolgerung, dass Bewertungen männlicher Sprecher, die auf deren SG basieren, in den USA und Korea gegensätzliche Tendenzen aufweisen und der kulturelle Hintergrund die auf der SG beruhende Wahrnehmung von Glaubwürdigkeit stark beeinflusst.¹⁹³ Neben dem oben beschriebenen Zusammenhang von SG und wahrgenommener Kompetenz wird eine weitere mögliche Erklärung angeführt: Da niedrige SG mit hohem Alter in Verbindung gebracht wird und Wissen und Kompetenz in der koreanischen Kultur ebenfalls mit fortgeschrittenem Alter assoziiert werden, kann so ein Zusammenhang zwischen niedriger SG indirekt mit wahrgenommener Kompetenz erklärt werden.

Hinweise auf unterschiedliche sprachliche Umsetzung von Emotionen in Abhängigkeit des kulturellen Hintergrundes der Sprecher legen u.a. Braun/Oba (2009) vor. 45 Szenen aus der Serie „Ally McBeal“ in der US-amerikanischen Originalfassung und der deutschen und japanischen Übersetzung, die von einem Publikum von zehn Personen bestimmten Emotionen zugeordnet wurden, werden auf Pausenstruktur, SG und AG sowie Häsitationsphänomene hin untersucht. Es fallen unterschiedliche Pausenlängenverteilungen gemäß den Emotionen auf: Die Sprecherin der japanischen Version produziert die längsten Pausen, daneben produziert der Sprecher der japanischen Version im Gegensatz zum deutschen und US-amerikanischen Sprecher in den Emotionen Ärger (allgemein), Angst und Freude kürzere Pausen im Vergleich mit emotional neutralen Situationen. Daneben sind deutliche Unterschiede bezüglich der AG/SG festzustellen: Die japanische Sprecherin zeichnet sich durch eine deutlich höhere SG bei Wut (spontan/momentan) und Ärger (allgemein) aus, wohingegen die anderen Sprecherinnen in diesen Emotionen langsamer als in neutralen Situationen sprechen. Unter den Sprechern fallen die niedrigeren Werte des deutschen Sprechers für Wut (spontan/momentan), Freude und Traurigkeit auf. In Bezug auf die AG sind für die japanische Sprecherin deutlichere Abweichungen von der emotional neutralen Sprechsituation festzuhalten, v.a.

¹⁹³ Andere Ergebnisse liefern Peng et al. (1993): Sie können keinen Zusammenhang von wahrgenommener Kompetenz und SG bei koreanischen Hörern finden, allerdings wird in den Aufnahmen das englische bzw. koreanische Alphabet vorgelesen, was eine außergewöhnliche Redeaufgabe darstellt und das Ergebnis beeinflussen kann. In den USA lebende Koreaner wiederum zeigen die oben beschriebenen Effekte (hohe SG = Kompetenz), so dass zumindest ein Unterschied zwischen in der koreanischen und in der US-amerikanischen Gesellschaft lebenden Individuen bestätigt werden kann.

spontane/momentane Wut wird von erhöhter AG begleitet, was bei den anderen beiden Sprecherinnen nicht der Fall ist. Der deutsche Sprecher artikuliert in allen Emotionen langsamer, wohingegen der japanische und der US-amerikanische Sprecher bei momentaner Wut schneller artikulieren. Stichprobenumfang und fehlende statistische Berechnungen lassen keine genauen Schlussfolgerungen zu, generell soll damit aber als nachweisbar gelten, dass gleiche Emotionen von unterschiedlichen sprachlichen (hier prosodischen) Erscheinungen begleitet werden können, sofern postuliert werden kann, dass eine Situation in verschiedenen Kulturen gleiche Reaktionen hervorruft - für übersetzte Filme/Serien soll davon ausgegangen werden, dass der emotionale Gehalt „mitübersetzt“ wird.

Auch auf einen unterschiedlichen Einsatz der SG zum Ausdruck von Höflichkeit liegen Hinweise vor (z.B. in Grawunder/Winter, 2010). Koreanische Sprecher (N=16) produzieren in höflichen Äußerungen (z.B. gegenüber einem Professor im Vergleich zu einem Kommilitonen) signifikant mehr Häsitationen/Verzögerungen und Atempausen, was für deutsche Sprecher (N=9) nicht nachweisbar ist. Die AG wird dabei nicht verändert. Daneben weist Lebra (1987) auf lautliche Verzögerungen in Zusammenhang mit Höflichkeit (u.a.) bei japanischen Sprechern hin.

Unter Berücksichtigung der genannten Studien müssen die Ergebnisse der US-amerikanischen Persönlichkeitswahrnehmungsforschung unbedingt auf den jeweils untersuchten kulturellen Kontext eingeschränkt werden.¹⁹⁴ Eine Übertragung von Assoziationen wie z.B. der hohen Sprechtempos mit Kompetenz kann zu erheblichen Missverständnissen in interkulturellen Kommunikationssituationen führen.

6.4 Vergleich der Ergebnisse aus Kapitel 4 mit kulturellen Dimensionen

Die umfassendste verhaltensorientierte empirische Untersuchung zu kulturellen Unterschieden ist die GLOBE-Studie (House et al., 2004). In ihr wurden bisher über 17.000 Führungskräfte des mittleren Managements verschiedener Branchen aus 62 Gesellschaften¹⁹⁵ mittels Fragebögen zu ihrem (hauptsächlich wirtschaftlichen) Organisations- und Führungsstil befragt. Neun kulturelle Dimensionen konnten festgestellt

¹⁹⁴ Weitere Studien zum Vertrauen gegenüber dem Gesprächspartner zeigen ähnliche Verhältnisse, so fördert hohe SG das Vertrauen in den USA (Scherer, 1979a), nicht jedoch im Iran (Zandpour/Sadry 1996); nach Traunmüller (2004: 659).

¹⁹⁵ Die Studie wurde bisher in 59 Ländern durchgeführt, wobei 62 Kulturen differenziert werden; in besonders heterogenen Gesellschaften wie der US-amerikanischen liegt jedoch keine Unterteilung vor.

werden, anhand derer sich Unterschiede erklären lassen:¹⁹⁶ *Uncertainty Avoidance, Power Distance, Institutional Collectivism, In-Group Collectivism, Gender Egalitarianism, Assertiveness, Future Orientation, Humane Orientation* und *Performance Orientation* (relevante Definitionen s.u.). Obwohl Manager aufgrund ihrer im Zuge der Globalisierung angeeigneten interkulturellen Kompetenz vermutlich nicht die ideale Population darstellen, um Kulturen zu untersuchen, sollen die Daten der Studie zum Vergleich herangezogen werden.

Um zu prüfen, ob neben phonetisch-phonologischen Variablen kulturelle Unterschiede verschiedene Sprechtempi erklären können, wurden die Ergebnisse aus Kapitel 4 mit den entsprechenden Werten aus der GLOBE-Studie auf Korrelationen hin untersucht. Dabei wurden alle Aspekte des Sprechtempos (SGs und AGs jeweils in Lauten und Silben/s sowie Wort- und Morphemgeschwindigkeit) berücksichtigt. In Tabelle 6.1 werden nur Fälle dargestellt, in denen mehrere Aspekte des Tempos (annähernd) signifikant miteinander korrelieren.¹⁹⁷

¹⁹⁶ Es handelt sich um eine Erweiterung des vier-Dimensionen-Modells nach Hofstede (1980). Ursprünglich untersuchte Hofstede Verhalten und Denkweisen von über 100.000 IBM-Mitarbeitern in 60 Ländern und konnte mit seinen Daten vier kulturelle Dimensionen aufstellen. Die Studie ist am Einfluss des kulturellen Hintergrundes auf den wirtschaftlichen Führungsstil interessiert und wurde seitdem mehrmals erweitert. In der GLOBE-Studie wird außerdem zwischen praktizierten und angestrebten Werten unterschieden (s.u.).

¹⁹⁷ Alle Werte aus der GLOBE-Studie liegen zwischen 1 und 7, da es sich um die gemittelten Werte aus Fragebögen mit 7-Punkte-Skalen handelt. Zu allen vollständigen Definitionen von kulturellen Dimensionen s. House et al. (1999).

Tabelle 6.1: Korrelationen zwischen Sprechtempo und kulturellen Dimensionen (Ist- und Soll-Werte)

Tempomaß		Kulturelle Dimension					
		Future Orientation	FO should be	Collectivism (Group) ¹⁹⁸	COL(Group) should be	Uncertainty Avoidance	UA should be
Laute/s (AG)	r	-0,462	-0,304	0,571	-0,439	0,182	0,732
	p	0,357	0,558	0,237	0,384	0,730	0,098
Silben/s (AG)	r	-0,852*	0,770	0,661	0,254	-0,801	0,655
	p	0,031	0,073	0,153	0,628	0,055	0,158
Laute/s (SG)	r	-0,420	-0,302	0,643	-0,429	0,039	0,877*
	p	0,407	0,561	0,168	0,396	0,942	0,022
Silben/s (SG)	r	-0,838*	0,727	0,721	0,141	-0,820*	0,753
	p	0,037	0,101	0,106	0,790	0,046	0,084
Wörter/s	r	-0,433	0,789	0,254	0,579	-0,876*	0,374
	p	0,391	0,062	0,628	0,229	0,022	0,465
Morpheme/s	r	-0,746	-0,028	0,945**	-0,152	-0,264	0,924**
	p	0,089	0,958	0,004	0,774	0,613	0,008

Mit den Werten für *Future Orientation* (FO) korreliert die Silbengeschwindigkeit (AG und SG) stark negativ (signifikant), auch für die Morphemgeschwindigkeiten ergeben sich starke negative Korrelationen (nicht signifikant) und die Lautgeschwindigkeiten korrelieren mittel (negativ, nicht signifikant). FO wird definiert als „the degree to which a collectivity encourages and rewards future-oriented behaviors such as planning and delaying gratification“ (House et al., 2004: 282, nach House et al., 1999). Kulturen unterscheiden sich z.B. darin, welche Konzepte von Zeit zugrunde liegen (polychrone vs. monochrone Kulturen, s. 7.3.1). Ein Einfluss von FO auf das Sprechtempo ist insofern schwer nachvollziehbar, als dass es sich bei dieser kulturellen Dimension um ein langfristig orientiertes Denken handelt, das in keinem Zusammenhang mit der momentanen Verfassung und (Sprech-)Situation stehen muss. House et al. (2004: 309) machen darauf aufmerksam, dass FO positiv mit *Uncertainty Avoidance* (UA) korreliert. Diese bezieht sich u.a. auf

„the extent to which ambiguous situations are threatening to individuals, to which rules and order are preferred, and to which uncertainty is tolerated in a society [...], to which members of collectives seek orderliness, consistency, structure, formalized procedures, and laws to cover situations in their daily lives.“ (House et al. 2004: 602 f.)

¹⁹⁸ In Abgrenzung zu Individualismus/Kollektivismus auf der institutionellen Ebene. In der GLOBE-Studie wird zwar „Individualism“ als Bezeichnung für diese Dimension verwendet, da höhere Werte aber für stärker kollektivistisch geprägtes Verhalten stehen, soll sie hier als „Collectivism“ bezeichnet werden.

Die kulturelle Dimension „Unsicherheitsvermeidung“ beschreibt eine Einstellung, die durchaus Einfluss auf tägliche Verhaltensweisen haben kann. Darüber hinaus muss in Erinnerung gerufen werden, dass das untersuchte Sprachmaterial in formellen, normierten Situationen produziert wurde. Dass hier in Kulturen, die zur Unsicherheitsvermeidung und damit starker Reglementierung neigen, Einschränkungen bezüglich der Schnelligkeit der Rede bestehen, ist nachvollziehbar.¹⁹⁹ Silben- und Wortgeschwindigkeit korrelieren signifikant stark negativ mit den UA-Werten, die Soll-Werte (gewünschte Unsicherheitsvermeidung) korrelieren stark negativ (signifikant) mit der Lautgeschwindigkeit und stark negativ (hochsignifikant) mit der Morphemgeschwindigkeit. Es ist damit nicht zu gewagt, festzuhalten, dass in Kulturen, in denen unsicherheitsvermeidendes Verhalten gefördert wird, in gewissen Sprechsituationen Tendenzen zu niedrigem Sprechtempo bestehen.

In Levine/Norenzayan (1999) wird ein Zusammenhang höherer Geschwindigkeiten bei Verhaltensweisen im täglichen Leben mit Individualismus festgestellt, der allerdings über die wirtschaftliche Entwicklung erklärt werden kann: Wirtschaftlich weiter entwickelte Gesellschaften sind in der Regel stärker von Individualismus geprägt, legen größeren Wert auf Leistung und neigen zu einer strikteren Handhabung von Zeitplänen und organisierten Abläufen; Zeit wird tendenziell als objektive Größe empfunden. Da kollektivistisch geprägte Kulturen zu HC-Kommunikation neigen (House et al. 2004: 454) und diese oben als indirekt und damit potentiell langsamer beschrieben wurde, gilt auch in diesem Sinne ein höheres Sprechtempo in stärker individualistisch geprägten Kulturen als nachvollziehbar. Die Korrelationen der GLOBE-Werte mit den Ergebnissen aus Kapitel 4 zeigen jedoch ein konträres Bild: Zwischen Kollektivismus und Tempo bestehen mittlere bis hohe positive Korrelationen, eine hochsignifikante positive Korrelation liegt mit der Morphemgeschwindigkeit vor, d.h. geringerer Individualismus geht mit höherem Sprechtempo bzw. höherer Informationsdichte (im morphologischen Sinn) einher. Für dieses Ergebnis kommen zwei Erklärungsansätze in Frage. Zum einen ist es möglich, dass die untersuchte Sprechsituation einen Sonderfall darstellt und nicht von dem Zusammenhang zwischen Individualismus und höherem Tempo erfasst wird, der nur auf „normale“ Handlungen zutrifft. Da vorbereitetes Sprechen in einer formellen Situation von nicht durch diaphasisch hoch markierte Redeweisen unterschieden werden muss, gelten hier andere Grundsätze, nach denen in individualistisch geprägten Kulturen niedriges, in

¹⁹⁹ Ob diese in Nachrichtensendungen vom Sender vorgegeben sind, sei dahingestellt. Wie bereits erwähnt ist eine Vorgabe für Laut-, Silben- und Morphemgeschwindigkeit unwahrscheinlich und kann in den untersuchten deutschen und brasilianischen Nachrichtensendungen ausgeschlossen werden.

kollektivistisch geprägten Kulturen hohes Sprechtempo gefordert wird. Der in den Nachrichtensendungen vorliegende Kontext muss jedoch enger gefasst werden, da es unwahrscheinlich ist, dass z.B. politische Reden, die auch als formell gelten, auch von derartigen Regeln erfasst werden.²⁰⁰ Konkreter können diese Grundsätze für informative Texte in Medien postuliert werden. Dabei ergibt sich das Problem, dass unter Berücksichtigung der im vorigen Abschnitt erwähnten Studien diese informativen Texte in Medien aufgrund des niedrigeren Sprechtempos in individualistisch geprägten Kulturen mit negativen Bewertungen bezüglich Glaubwürdigkeit, Überzeugungskraft, Sachkenntnis usw. einhergehen und in kollektivistisch geprägten Kulturen das höhere Tempo ebenso – zumindest teilweise – mit negativen Urteilen zusammenhängt.²⁰¹ Überzeugender ist der folgende Erklärungsansatz: Da bekanntlich verschiedene Verhaltensweisen (zumindest in Bezug auf den Ausdruck von Emotionen) von Individuen gegenüber anderen Individuen nachweisbar sind, in Abhängigkeit davon, ob Letztere der Eigengruppe oder der Fremdgruppe zugerechnet werden (Matsumoto et al. 2008), liegt nahe, dass in den untersuchten Kulturen unterschiedlich zwischen Eigen- und Fremdgruppe differenziert wird.²⁰² Je nachdem, ob der Empfänger der Nachricht als Teil der Eigen- oder Fremdgruppe erachtet wird, wird ein höheres oder niedrigeres Sprechtempo gewählt. In individualistisch geprägten Kulturen geht dann die Textsorte „informativer Text in den Medien“ mit einer Wertung der Empfänger als Teil der Fremdgruppe einher, was ein niedriges Sprechtempo erfordert. In kollektivistisch geprägten Kulturen dagegen erfordert der gleiche Kontext die Anwendung von Regeln für die Eigengruppe, die ein höheres Sprechtempo begünstigen. Die bei niedrigem Sprechtempo wahrgenommenen negativen Eigenschaften (s.o.) können dann auf den Kontext „Kommunikation in der Eigengruppe in

²⁰⁰ Intuitiv betrachtet und dem subjektiven Höreindruck nach zu urteilen liegt in politischen Reden in kollektivistisch geprägten Kulturen niedriges Sprechtempo vor.

²⁰¹ Natürlich kann sich die Wahrnehmung des Tempos an den situativen Kontext anpassen. Hörer verwenden dann in Abhängigkeit der Sprechsituation unterschiedliche Maßstäbe zur Beurteilung von Kompetenz, usw. Vgl. hierzu Laver (1994: 544), der Untersuchungen zusammenfasst, die keinen direkten Zusammenhang von Formalität und Sprechtempo vermuten lassen, im Gegensatz zu Überlegungen in Monroy Casas (2008).

²⁰² In Matsumoto et al. (2008) wird auch beschrieben, dass Eigen- und Fremdgruppe in individualistisch bzw. kollektivistisch geprägten Gesellschaften unterschiedliche Bedeutungen haben. In IDV-Kulturen gehören Individuen mehr Eigengruppen an, die einen geringeren Grad an Kohäsion aufweisen. In der (Sozial-) Psychologie gehen die Begriffe Eigen- und Fremdgruppe für Gewöhnlich mit einer tendenziellen Abwertung der Fremd- und einer tendenziellen Aufwertung der Eigengruppe einher, zumindest aber mit stereotypem bzw. vorurteilsbehaftetem Denken („Fremdgruppenhomogenität“). Es soll an dieser Stelle nachdrücklich darauf hingewiesen werden, dass derartige Konnotationen hier nicht intendiert sind, sondern eine Identifikation und direkte Verbindung im Kommunikationsprozess zwischen Nachrichtensprecher und Zuschauer gemeint ist. Dieses „Wir-Gefühl“ kann durchaus vorsätzlich vermittelt, aber vom Nachrichtensprecher als Individuum nicht empfunden sein.

individualistisch geprägten Kulturen“ eingeschränkt werden.²⁰³ In diesem Sinne lehnt sich die Unterscheidung der kulturellen Systeme an eine in expressive und reservierte Kulturen an.²⁰⁴

Vorläufig soll festgehalten werden, dass das Sprechtempo mit kulturellen Merkmalen korreliert, die in verschiedenen situativen Kontexten unterschiedliche Auswirkungen auf das sprachliche Verhalten haben können. In Kapitel 8 wird darauf eingegangen, ob die Untersuchung der Verwendung nonverbaler Kommunikationsmittel in den Nachrichtensendungen den zweiten Erklärungsansatz untermauert. Dabei wird auch auf die kulturelle Dimension *Gender Egalitarianism* eingegangen, die an dieser Stelle nicht behandelt werden kann.

6.5 Umfrage zu kulturellen Stereotypen

Im Rahmen einer eigenen Erhebung konnten in 5.4.5 erwähnte Stereotype bestätigt werden: Mehr als 86% von 51 befragten Romanistikstudenten geben an, dass Sprachen nicht gleich schnell gesprochen werden und fast 75% stimmen der Aussage zu, dass BP und Spanisch schneller gesprochen werden als Deutsch und Englisch.²⁰⁵ Im Fragebogen wurde die Aussage „Fremdsprachen klingen schneller als die Muttersprache“ eingeschlossen, um die Teilnehmer auf ihre subjektive, als Lerner einer oder mehrerer Fremdsprachen möglicherweise voreingenommene Wahrnehmung aufmerksam zu machen. Mehr als zwei Drittel der Befragten stimmen dieser Aussage zu. In Fragen zum direkten Vergleich der SG von Völkern²⁰⁶ geben jeweils rund 85 bzw. über 86% an, Briten sprechen langsamer als Spanier und diese schneller als Deutsche. Weniger Übereinstimmung besteht im Vergleich fremder Nationalitäten untereinander. Brasilianer werden von 47% der Befragten als

²⁰³ Obwohl die situativen Gegebenheiten in den oben erwähnten Untersuchungen nicht identisch sind, kann zumindest für die Untersuchungen, in denen Passanten auf der Straße von vermeintlichen Radiojournalisten angesprochen werden, angenommen werden, dass die Teilnehmer sich als mit den Befragenden in der Eigengruppe befindlich empfinden. Gleiches gilt für Untersuchungen, die in Kursen an Universitäten durchgeführt werden (in denen davon ausgegangen werden kann, dass die Studenten einen gewissen Grad an Vertrautheit miteinander empfinden). Diese Überlegungen sind jedoch spekulativ und müssen experimentell unter Berücksichtigung individueller Unterschiede überprüft werden.

²⁰⁴ Vgl. Wierzbickas Beschreibung kultureller Skripte zur Expressivität in 7.3.3. Allgemeine emotionale Expressivität und Individualismus wurden allerdings auch in Zusammenhang gebracht (s. Matsumoto et al., 2008), was Überlegungen Wierzbickas widerspricht und weiter unten angesprochen wird.

²⁰⁵ Die Umfrage wurde an der Friedrich-Alexander-Universität Erlangen-Nürnberg durchgeführt, wo BP nicht als Studienfach angeboten wird.

²⁰⁶ Die Formulierung im Fragebogen (s.u.) wurde gewählt, weil kulturelle Klischees Gegenstand der Untersuchung sind und um eine linguistische Betrachtung möglichst auszuschließen (Bsp.: „Spanier sprechen...“, mögliche Antworten: „... genauso schnell wie Briten / schneller als Briten / langsamer als Briten / weiß ich nicht“).

langsamere Sprecher im Vergleich mit Spaniern eingeschätzt. Dass bei einem Vergleich von Chinesen und Deutschen tendenziell Chinesen als die schnelleren Sprecher bezeichnet werden (53%) weist darauf hin, dass fremde (unvertraute) Sprachen generell mit höherer SG assoziiert werden. In einem Teil der Befragung, der frei zu beantworten war, zeichnet sich eine Tendenz ab, romanische Sprachen mit hoher, nordeuropäische mit niedriger SG zu assoziieren. Die Befragung wurde mit teilweise leicht abgeänderten Fragen in zwei weiteren Gruppen durchgeführt: In China lebende Chinesen (N=56), die ein nicht-sprachliches Fach an einer Hochschule in Qingdao studieren, und in Deutschland lebende Chinesen, die sich im Anschluss an ihr Studium auf eine Tätigkeit als Chinesischlehrer vorbereiten (N=14). In beiden Gruppen lautet beim Vergleich von Chinesen mit Deutschen die häufigste Antwort, dass Chinesen langsamer sprechen als Deutsche (79% bzw. 57% unter den in Deutschland lebenden Chinesen). Auch bei einer Gegenüberstellung von westlichen Sprachen mit Chinesisch werden die westlichen Sprachen von 68% der in China lebenden Chinesen als schneller bezeichnet, allerdings geben nur rund 29% der in Deutschland lebenden Chinesen diese Antwort und 32% der in China lebenden Chinesen betonen, dass westliche Sprachen nicht einheitlich schnell gesprochen werden. 66% der in China lebenden Chinesen stimmen der Aussage, Deutsch und Englisch werden langsamer gesprochen als BP und Spanisch, nicht zu; von den in Deutschland lebenden Chinesen stimmen dagegen rund 79% zu. Dass Fremdsprachen schneller klingen als die Muttersprache wird von 91% der in China lebenden Chinesen bestätigt, wohingegen von den Chinesen in Deutschland nur 57% zustimmen. In Tabelle 6.2 werden die Fragen, die in den Fragebögen der drei Gruppen enthalten sind, und die häufigsten Antworten in Prozent angegeben.

Einige Antworten können auf den Grad der Vertrautheit mit der jeweiligen Fremdsprache zurückgeführt werden (z.B. die Antworten chinesischer Befragter zum Vergleich britischer mit deutschen Sprechern, da viele der Teilnehmer Englisch gelernt hatten, aber nicht oder kaum Deutsch). Erwähnenswert ist die Tatsache, dass in Deutschland lebende Chinesen, die einen größeren Erfahrungsschatz im Umgang mit Fremdsprachen haben, seltener angeben, dass Fremdsprachen schneller klingen als die Muttersprache, was möglicherweise auf einen höheren Grad an Kenntnissen in zumindest einer Fremdsprache zurückgeführt werden kann. Ohne auf weitere Details einzugehen kann hier festgehalten werden, dass unabhängig davon eine Tendenz besteht, fremde Sprachen als schneller zu empfinden (bei Präsentation mit Stimuli) und zu vermuten (ohne Präsentation). Ebenso werden einzelne romanische Sprachen in allen Gruppen tendenziell als schneller im Vergleich mit bestimmten germanischen Sprachen bezeichnet (vgl. 2.3.2, 4.5.7.1 und 5.5.1).

Tabelle 6.2: Ergebnisse der Befragungen zur typischen SG anderer Nationalitäten

Frage / Aussage	Gruppe		
	Deutsche	Chinesen (in China)	Chinesen (in Deutschland)
	Häufigste Antwort / Beurteilung:		
„Fremdsprachen klingen schneller als die Muttersprache.“	Stimme zu 67%	Stimme zu 91%	Stimme zu 57%
„Alle genannten Sprachen werden gleich schnell gesprochen.“	Stimme nicht zu 86%	Stimme nicht zu 98%	Stimme zu 36% Stimme nicht zu 36%
„Englisch und Deutsch werden langsamer gesprochen als BP und Spanisch.“	Stimme zu 75%	Stimme nicht zu 66%	Stimme zu 79%
CN und DE im Vergleich	CN schneller 53%	CN langsamer 78%	CN langsamer 57%
ES und DE im Vergleich	ES schneller 86%	ES schneller 39%	Weiß nicht 36% ES schneller 29% Gleich schnell 29%
BP und ES im Vergleich	BP langsamer 47%	Gleich schnell 29%	Weiß nicht 64%
EN und DE im Vergleich	Gleich schnell 53%	EN langsamer 66%	EN langsamer 57%
Westliche Sprachen und CN im Vergleich	Westl. untereinander nicht einheitlich 57%	Westl. schneller 68%	Westl. untereinander nicht einheitlich 36%

6.6 Wahrnehmungsexperiment zur Bewertung von Sachlichkeit und Vertrauenswürdigkeit

Als für Nachrichtensendungen relevante Beurteilungen durch das Publikum gelten u.a. „objektiv“ und „sachlich“, aber auch „verlässlich“, in Bezug auf die Quelle. Da die untersuchten Nachrichtensendungen nicht nur was das Sprechtempo anbelangt Unterschiede aufweisen (s. Kap. 8) liegt die Vermutung nahe, dass die genannten Attribute in den untersuchten Kulturen unterschiedlich konstruiert werden und mit der Verwendung unterschiedlicher sprachlicher Mittel einhergehen. Um dies zu überprüfen wurde ein Versuch durchgeführt, bei dem chinesische und deutsche Versuchsteilnehmer Sachlichkeit und Vertrauenswürdigkeit in zehn kurzen Ausschnitten aus den untersuchten Nachrichtensendungen zu beurteilen hatten. Für dieses Experiment ist das untersuchte Material deshalb problematisch, weil es aufgrund des geringen Umfangs in Bezug auf die phonetische Beschaffenheit schwer kontrollierbar ist und nicht uneingeschränkt als repräsentativ für die

Einzel sprachen bezeichnet werden kann. Da aber erste Hinweise auf die spontanen Reaktionen und Beurteilungen der beiden Gruppen von Hörern Hauptgegenstand der Untersuchung sind, kann dieses unmanipulierte natürliche Sprachmaterial als angemessen bezeichnet werden.

6.6.1 Methode

Zu den Nachrichtensendungen wurden je zwei Fragen auf einem Fragebogen beantwortet. Der Aussage „Die Aufnahme lässt vermuten, dass der Inhalt objektiv und sachlich dargestellt wird“ konnte „voll und ganz“, „teilweise“ oder „nicht“ zugestimmt werden (folgend als „Sachlichkeit“ bezeichnet). Weiterhin konnte die Informationsquelle als „zuverlässig“ oder „nicht sehr zuverlässig“ bezeichnet werden (folgend als „Vertrauenswürdigkeit“ in Bezug auf die Aufnahme bezeichnet). Für diese Kategorien wurden bei der Auswertung absolute Werte vergeben und die arithmetischen Mittelwerte der Beurteilungen berechnet. Die Teilnehmer wurden aufgefordert, die Sprechweise in den Aufnahmen spontan und subjektiv zu beurteilen und abschließend um eine kurze schriftliche Beschreibung, worauf sie bei ihrer Beurteilung geachtet haben, gebeten. Über die Aufnahmen wurde den Teilnehmern nur mitgeteilt, dass sie den Inhalt der teilweise unverständlichen Aufnahmen außer Acht lassen sollten. Die Aufnahmen wurden allen Teilnehmern gleichzeitig in der gleichen Reihenfolge über Lautsprecher präsentiert, zwischen zwei Aufnahmen wurden zehn Sekunden Zeit für die Beurteilung gegeben. Die Übersetzung der Fragebögen ins Chinesische wurde von drei Muttersprachlern unabhängig voneinander kontrolliert.

6.6.2 Teilnehmer

56 chinesische Studenten einer nicht-sprachlichen Fachrichtung im ersten bis zweiten Semester wurden an ihrer Heimatuniversität in Qingdao zu dem Versuch in das örtliche Sprachlabor eingeladen. Die deutsche Gruppe bestand aus 68 Studenten verschiedener Fachrichtungen im ersten bis vierten Semester, die ebenfalls an ihrer Heimatuniversität zur Durchführung des Tests eingeladen wurden.

6.6.3 Stimuli

Die Aufnahmen von je einer Sprecherin und einem Sprecher der Sprachen BP, Chinesisch, Deutsch, Englisch und Spanisch betragen zwischen 15 und 17 s Länge und stellen einen unmanipulierten Ausschnitt aus den untersuchten Nachrichtensendungen dar. Alle

Aufnahmen umfassen mindestens einen syntaktisch vollständigen, unabhängigen Satz, der am Ende der Aufnahme syntaktisch und inhaltlich vollständig und intonatorisch als beendet markiert ist. Da Tempo, f₀-Verlauf und Pausenstruktur unverändert blieben, ergeben sich Unterschiede, die in Tabelle 6.3 dargestellt sind.

Tabelle 6.3: Beschaffenheit der Aufnahmen im Versuch zur wahrgenommenen Sachlichkeit und Vertrauenswürdigkeit

Aufnahme	Dauer (s)	Tempo (AG)		Tonhöhenwechsel (ca. 50 Hz)	Pausen
		Laute/s	Silben/s		
BPw	15,425	15,88	6,96	23	8
BPm	14,708	15,35	6,81	27	4
CNw	15,956	18,18	5,65	29	5
CNm	16,892	18,72	6,09	17	6
DEw	17,364	15,18	5,22	9	3
DEm	14,970	15,70	5,04	7	3
ENw	13,452	14,35	4,46	26	4
ENm	12,232	16,92	5,64	13	3
ESw	13,987	17,52	7,65	13	2
ESm	16,843	16,18	6,97	12	5

6.6.4 Ergebnisse und Diskussion

Die Aufnahmen BPw, CNw, CNm, DEw, DEm und ESw werden in Bezug auf Sachlichkeit oder Vertrauenswürdigkeit von chinesischen und deutschen Versuchsteilnehmern signifikant unterschiedlich bewertet (s. Tab. 6.4). Damit steht zunächst fest, dass diese Konzepte in gesprochenen Texten von chinesischen und deutschen Hörern unterschiedlich konstruiert werden. Im frei zu beantwortenden Abschnitt des Fragebogens geben 75% der chinesischen Studenten an, bei der Beurteilung auf das Sprechtempo geachtet zu haben, allerdings ohne weitere Angaben, ob ein hohes oder ein niedriges Tempo mit Sachlichkeit bzw. Vertrauenswürdigkeit in Verbindung gebracht wird. 57% erwähnen die Intonation und 13% den emotionalen Charakter²⁰⁷ der Aufnahmen. Unter den deutschen Versuchsteilnehmern lauten die häufigsten Antworten Sprechtempo (38%), Intonation (34%), artikulatorische Genauigkeit²⁰⁸ (24%) und Betonungen (22%). Der f₀-Verlauf wird dabei unterschiedlich bewertet: Teilweise wird eine dynamische, überwiegend jedoch eine „emotionslose“ Intonation mit Objektivität in Verbindung

²⁰⁷ „感情“, de. „Emotion“, wobei damit nicht der f₀-Verlauf gemeint sein muss, den einige Teilnehmer zusätzlich erwähnen.

²⁰⁸ Bezeichnet als „Deutlichkeit“, „Klarheit“ oder „Aussprache“.

gebracht. Bezüglich des Sprechtempos wird niedriges, aber auch konstantes Tempo als sachlich bezeichnet.

Tabelle 6.4: Signifikante Unterschiede in der Beurteilung von Sachlichkeit bzw. Vertrauenswürdigkeit der einzelnen Ausschnitte aus Nachrichtensendungen durch chinesische (N=56) und deutsche (N=68) Hörer; rechts die Gruppe, für die auf den jeweiligen Aspekt bezogen ein höherer gemittelter Wert vorliegt

Aufnahme	Aspekt	T	df	p	Höher eingeschätzt von
BPw	Sachlichkeit	3,568	108	0,001	deutschen Teilnehmern
BPw	Vertrauenswürdigkeit	2,654	121	0,009	deutschen Teilnehmern
CNw	Sachlichkeit	-7,338	105	<0,001	chinesischen Teilnehmern
CNw	Vertrauenswürdigkeit	-3,740	120	0,001	chinesischen Teilnehmern
CNm	Sachlichkeit	-1,608	120	0,001	chinesischen Teilnehmern
DEw	Sachlichkeit	3,546	119	0,001	deutschen Teilnehmern
DEm	Sachlichkeit	4,569	122	<0,001	deutschen Teilnehmern
DEm	Vertrauenswürdigkeit	1,945	122	0,055	deutschen Teilnehmern
ESw	Sachlichkeit	3,335	111	0,001	deutschen Teilnehmern
ESw	Vertrauenswürdigkeit	3,933	121	<0,001	deutschen Teilnehmern

Aus einer Korrelationsmatrix (s. Tab. 6.5) geht hervor, dass die Beurteilung der Sachlichkeit durch deutsche Studenten (gemittelte Werte) negativ mit der Anzahl der Tonhöhenwechsel (50 Hz; $r=-0,655^{**}$) und der Anzahl der Pausen ($r=-0,660^{*}$) korreliert. Die Beurteilung der Vertrauenswürdigkeit korreliert mit den gleichen Eigenschaften (Tonhöhenwechsel: $r=-0,637^{*}$; Pausen: $r=-0,67^{*}$), wahrgenommene Sachlichkeit und Vertrauenswürdigkeit korrelieren stark miteinander ($r=0,979^{**}$). Für die chinesische Gruppe ergeben sich Korrelationen nur mit der Silbengeschwindigkeit (Sachlichkeit: $r=-0,681^{*}$; Vertrauenswürdigkeit: $r=-0,765^{**}$), Sachlichkeit und Vertrauenswürdigkeit korrelieren ebenfalls stark positiv miteinander ($r=0,906^{**}$). Deutsche Hörer scheinen dem f0-Verlauf und der SG bzw. Pausenstruktur besondere Aufmerksamkeit zu schenken, wenn Aufnahmen auf die genannten Attribute hin beurteilt werden, was im Allgemeinen den Kommentaren der Versuchsteilnehmer selbst entspricht. Für chinesische Hörer dagegen scheint die *objektiv gemessene* AG (Silbengeschwindigkeit) eine größere Rolle zu spielen. Zu erwarten war die Einschätzung der Aufnahmen in der Muttersprache der Versuchsteilnehmer. Diese Aufnahmen wurden fast ausnahmslos durch die Muttersprachler selbst als sachlicher und vertrauenswürdiger bewertet.

Tabelle 6.5: Zusammenhänge zwischen gemittelter wahrgenommener Sachlichkeit und Vertrauenswürdigkeit chinesisch und deutscher Hörer mit Sprechtempomesswerten, f0-Wechseln und Pausenhäufigkeit

Prädikat		Tempo		Tonhöhenwechsel	Pausenhäufigkeit
		Laute/s	Silben/s		
Sachlichkeit (de)	r	0,360	0,588	0,655*	0,660*
	p	0,307	0,074	0,040	0,038
Vertrauenswürdigkeit (de)	r	0,265	0,487	0,637*	0,670*
	p	0,459	0,154	0,047	0,034
Sachlichkeit (cn)	r	-0,278	0,681*	0,004	0,279
	p	0,438	0,030	0,991	0,435
Vertrauenswürdigkeit (cn)	r	0,278	0,765**	0,185	0,293
	p	0,437	0,010	0,608	0,412

Obwohl die Übersetzung des Fragebogens im Detail mit Muttersprachlern besprochen wurde, kann nicht ausgeschlossen werden, dass die Versuchsteilnehmer (auch innerhalb der Gruppen) unterschiedliche Vorstellungen der zu bewertenden Aspekte haben und hierdurch verschiedene Bewertungen ausgelöst werden. Die Ergebnisse des Experiments müssen auch aufgrund des geringen Stichprobenumfangs mit besonderer Vorsicht interpretiert werden. Es darf festgehalten werden, dass das Sprechtempo einen Einfluss auf die wahrgenommene Sachlichkeit und Verlässlichkeit der Quelle hat und dieser in verschiedenen Gruppen unterschiedlich stark ausgeprägt ist. Im beschriebenen Experiment wurde nur der Einfluss der objektiv gemessenen AG untersucht, weshalb weitere Versuche unter Berücksichtigung des wahrgenommenen Tempos unbedingt notwendig sind.

Abschließend soll noch der Einfluss von Vorurteilen auf die Beurteilung fremdsprachlicher Aufnahmen unterstrichen werden, der aus den Kommentaren in den Fragebögen hervorgeht. Manche Teilnehmer erwähnen die „Sprache“, „Sprachgruppe“ und das „Herkunftsland“ der Aufnahmen ohne weitere Spezifizierung als Information, die bei der Beurteilung berücksichtigt wurde; auch, wenn sie die Sprache nicht eindeutig identifizieren konnten.²⁰⁹

6.7 Zusammenfassung

In diesem Kapitel wurde das Sprechtempo unter einem kulturellen Aspekt betrachtet. Dabei wurde deutlich, dass verschiedene SGs in Gesellschaften unterschiedlich bewertet werden, wahrgenommene Kompetenz z.B. nicht universell mit hohem Sprechtempo

²⁰⁹ In einem Fragebogen wurde gar „Misstrauen gegenüber Asiaten“ erwähnt.

einhergeht. Als Einflussfaktoren werden generell kulturelle Unterschiede angenommen, die anhand des ursprünglich von Hofstede (1980) erarbeiteten und u.a. von House et al. (2004) weiter entwickelten Modells kultureller Dimensionen festgemacht werden können. In individualistisch bzw. kollektivistisch geprägten Gesellschaften wird das Sprechtempo unterschiedlich bewertet, wobei eine Interaktion mit der Differenzierung der Eigen- von der Fremdgruppe wahrscheinlich ist. Ein Vergleich der in der Untersuchung von Nachrichtensendungen erhobenen Werte mit den Werten aus der GLOBE-Studie zeigt einen Zusammenhang niedrigen Sprechtempos mit Individualismus, hoher Unsicherheitsvermeidung und langfristig orientiertem Planen. Neben diesen Zusammenhängen, die vorläufig nur für Nachrichtensprache gelten, wurde die Bedeutung von Stereotypen und Vorurteilen betont. Umfragen und Versuche zeigen, dass Hörer eine Vorstellung vom Sprechtempo fremder, teilweise unbekannter Sprachen haben, die auf Stereotypen bezüglich der entsprechenden Nationalitäten aufbauen müssen. Bei der Beurteilung von Sachlichkeit und Vertrauenswürdigkeit der Nachrichtenmeldungen anhand der Sprechweise werden unterschiedliche Informationen im Sprachsignal beachtet, derer sich die Hörer bewusst sind. Die gleichen Aufnahmen können daher sehr unterschiedlich bewertet werden, da, wie im Versuch, für chinesische Hörer das Sprechtempo, für deutsche Hörer der Tonhöhenverlauf und die Pausenstruktur von Bedeutung sind.

Ergänzend soll hier mit aller Nachdrücklichkeit darauf hingewiesen werden, dass, wie u.a. von Nakane (2007: 23) kritisiert wird, viele Beschreibungen z.B. des Umgangs mit Schweigen in anderen Sprachen oder Kulturen auf persönlichen Erfahrungen beruhen.²¹⁰ Mangelndes Verständnis für die situativen Kontexte und Angemessenheit von Schweigen in fremden Kulturen führt zu falschen Einschätzungen, Stereotypen und Vorurteilen, weshalb ohne wissenschaftliche empirische Untersuchungen als Grundlage keine Schlüsse über den Umgang mit Pausen und auch dem Sprechtempo gezogen werden dürfen.

Anschließend wird die Schnelligkeit gesprochener Sprache im Bereich nonverbaler Kommunikation untersucht.

²¹⁰ Nakane bezieht sich auf Lebra (1987), diese Kritik trifft aber auch auf die zitierten Texte von Mehrabian und Hall zu.

7 Nonverbale Kommunikation im Sprachvergleich

7.1 Einleitung

In den letzten Kapiteln hat sich herausgestellt, dass ein vokaler sprachbegleitender Aspekt kommunikativen Verhaltens, das Sprechtempo, als kultur- oder sprachgemeinschaftsabhängig bezeichnet werden muss. An dieser Stelle wird ein weiterer relevanter Bereich sprachlicher Kommunikation betrachtet: die nonverbale Kommunikation. Nach einer einleitenden Definition und ergänzenden Beschreibung der Bedeutung nonverbaler Kommunikation für gesprochene Sprachen werden die einzelnen Teilbereiche und Klassifizierungen nonverbalen Verhaltens vorgestellt, wobei die Ausführungen auf grundlegende Begriffe und Konzepte eingeschränkt werden. Anschließend sollen Unterschiede zwischen Sprachgemeinschaften beschrieben werden, die im interkulturellen Vergleich der Verwendung nonverbaler Kommunikationsmittel festgestellt werden konnten. Im Vordergrund stehen dabei Gestik und Mimik. Am Ende des vierten Abschnitts wird auch auf einen Zusammenhang mit den bisherigen Ergebnissen und Erklärungsansätze eingegangen. Die Zusammenfassung leitet zu einer weiteren empirischen Untersuchung im achten Kapitel über, bei der wie im ersten Teil der Arbeit produzierte sprachliche Elemente pro Zeiteinheit gezählt und verschiedene Sprachen miteinander verglichen werden.

7.2 Definition und Bedeutung nonverbaler Kommunikation

7.2.1 Definition

Unter dem Begriff „nonverbale Kommunikation“ werden unterschiedliche Aspekte zusammengefasst. Wie Nöth (2002: 293) erklärt, umfasst der Begriff im weitesten Sinne „Kommunikation minus Sprache“;²¹¹ sprachbegleitendes Verhalten wird dagegen in einem engen Sinn darunter verstanden. Als Teilbereiche nonverbaler Kommunikation gelten Gestik, Kinesik, Mimik, Blickverhalten (Miremik²¹²), taktile Kommunikation (Haptik), Proxemik, Körperhaltung und Chronemik, teilweise auch Olfaktorik. Darüber hinaus

²¹¹ Unter „Sprache“ ist hier gesprochene Sprache zu verstehen, also die artikulatorischen Vorgänge.

²¹² Auch „Okulesik“ (Oksaar 1988: 37).

zählen auch schriftliche, bildliche (Künste) und artefaktische (z.B. Kleidung) Zeichensysteme zur nonverbalen und gleichzeitig nonvokalen Kommunikation (vgl. Kalverkämper 2003: 313 f.). Auch Phänomene, die unter dem Begriff „Paralinguistik“²¹³ behandelt werden, sind als nonverbale Zeichen, die sprachbegleitend verwendet werden, zu verstehen (z.B. Pausenverhalten). Eine scharfe Trennung zwischen vokaler nonverbaler Kommunikation (z.B. Fülllaute ohne lexikalischen Gehalt) und nonvokaler nonverbaler Kommunikation (z.B. redeersetzende Gesten) wird nicht immer gemacht.²¹⁴ Wie die Definition dessen, was mit dem Begriff „nonverbale Kommunikation“ bezeichnet wird, zeigt, ist der Bereich mindestens so facettenreich wie der verbaler Kommunikation. Deshalb und um eine negative Definition zu vermeiden werden oft alternative Ausdrücke, wie z.B. „Körpersprache“ vorgezogen. Im Folgenden wird auf die Aspekte Gestik, Mimik, Körperhaltung und Veränderungen derselben sowie Bewegungen des Kopfes während der Sprachproduktion eingegangen.

7.2.2 Bedeutung nonverbaler Kommunikation

7.2.2.1 Anteil nonverbaler Kommunikation an sprachlicher Kommunikation

Scollon/Wong Scollon (2002: 15 f.) fassen Schätzungen zusammen, welcher Anteil an Bedeutung über nonverbale Kommunikation im täglichen Sprachgebrauch vermittelt wird und erwähnen Zahlen von 65% (Westbrook Eakins/Eakins 1982: 292) bis 93% (Mehrabian, 1972). Realistischerweise sollten diese Angaben unter Vorbehalt zur Kenntnis genommen werden, auch unter der Berücksichtigung der Kritik von Scollon/Wong Scollon (2002), dass Bedeutung und, wie bei Westbrook Eakins/Eakins (1982), „soziale Bedeutung“ nur schwer messbar und als Wert darstellbar sind. Sicherlich kann davon ausgegangen werden, dass in vielen kommunikativen Situationen des täglichen Lebens der Großteil einer Nachricht über nonverbale Kanäle vermittelt werden *kann*. Grammer (2004: 3472 f.) erklärt, dass v.a. in mehrdeutigen Situationen, d.h. wenn der Inhalt der verbalen Nachricht und der Inhalt nonverbaler Zeichen nicht übereinstimmen, dieser Effekt verstärkt zu beobachten ist. Welche dieser Informationen für Empfänger (stärker) von Bedeutung ist und mit mehr Aufmerksamkeit verarbeitet wird, wenn widersprüchliche verbale und nonverbale Zeichen gesendet werden, kann, wie Ostermeier (1986: 85) zu bedenken gibt, jedoch auch von der Persönlichkeit abhängen (der des Hörers, aber auch der des Sprechers

²¹³ Ehlich (in Glück, 2005), lehnt den Begriff aus nachvollziehbaren Gründen ab.

²¹⁴ Vgl. Mehrabian (1981: 2), der von „nonverbal behavior“ gegenüber „‘implicit’ aspects of speech“ spricht.

– und daneben selbstverständlich von situativen und inhaltlichen Aspekten). Darüber hinaus scheinen die nonverbalen Kommunikationskanäle unterschiedlich stark genutzt zu werden, wobei der visuelle Kanal am stärksten verwendet wird.²¹⁵ Wie Birdwhistell (1970) festhält, kommuniziert der Mensch höchst selten einkanalig, sondern bedient sich im Normalfall mehrerer kommunikativer Kanäle gleichzeitig.

7.2.2.2 Zusammenhang von Sprache und Gestik / Kinesik

Für die Sprachproduktion stellt sich die Frage, inwiefern verbale und nonverbale Codes unabhängig voneinander (ontogenetisch) entwickelt und verwendet werden. In einigen Arbeiten kann ein direkter Einfluss auf die Sprachproduktion festgestellt werden, wenn die Verwendung von Gestik eingeschränkt wird (Dobrogaev, 1931; Rauscher/Kraus/Chen, 1996). Dabei nehmen die Flüssigkeit der Rede sowie die Bandbreite des verwendeten Vokabulars ab, wenn z.B. die Arme des Sprechers fixiert werden. Kendon (1983) sieht aufgrund der Synchronisierung von Gestik und verbaler Sprachproduktion sowie der Organisation beider kommunikativer Verhaltensweisen in miteinander vergleichbaren Einheiten (phonologische „*Tone Units*“, die mit nonverbalen „*Gesture Units*“ synchron sind) in der Gestik ein zwar von der verbalen Sprachproduktion unabhängiges, aber eng mit ihr verbundenes kommunikatives Verhalten:

„[I]t is as if gesture and speech are available as two separate modes of representation and are coordinate because both are being guided by the same overall aim. That aim is to produce a pattern of action that will accomplish the representation of a meaning.“ (Kendon 1983: 20)

Diese als „Central Organizer Concept“ bekannte Auffassung wird von Beobachtungen untermauert, nach denen prosodische Eigenschaften (f_0 , Akzente) mit gestischen Zeichen synchronisiert sind (Condon/Ogston, 1966; Condon, 1976; s.u.).

In anderen Ansätzen²¹⁶ werden weniger enge Zusammenhänge zwischen Sprache und Gestik gesehen. Dittman (1962), aber auch Ekman/Friesen (1969, 1972) sehen gestisches Verhalten als Ausdruck der Erregung. Positive Erregung steigert in ihren Theorien die Verwendung von Gestik, insbesondere von sprachbegleitenden, betonenden Gesten, wohingegen negative Erregung die Gestik einschränken kann. Ebenfalls mit Erregung bringt Freedman (1977) die Gestik in Verbindung, wobei er davon ausgeht, dass bestimmte gestische Verhaltensweisen die Aufmerksamkeit bei Ablenkung auf den Sprach-

²¹⁵ Grammer verweist hierzu auf Posner et al. (1976); vgl. auch den Überblick in Luccio (2004).

²¹⁶ Dargestellt in Wallbott (2003a: 270).

produktionsprozess zurücklenken können. In einer Untersuchung von Christenfeld et al. (1991) wird festgestellt, dass die Häufigkeiten von Häsitationsmitteln und Gestik negativ miteinander korrelieren. Die Autoren vermuten, dass Gestik eingesetzt wird, wenn in der Sprachproduktion ein passender Begriff gefunden und damit ein sprachliches Ziel erreicht wurde. Gestik wäre damit im Gegensatz zum Ansatz von Freedman nicht als die Aufmerksamkeit steuernder Prozess, sondern als Nebenprodukt der erfolgreichen Sprachproduktion zu verstehen. Sie muss also nicht nur als Begleiterscheinung während der Suche nach Worten verstanden werden.

Oft wird auf einen Zusammenhang zwischen prosodischen Eigenschaften und nonverbalem Verhalten hingewiesen.²¹⁷ Die Bewegungsrichtung der Hände kann beim Gestikulieren parallel zur Intonation verlaufen. Condon (1976) erwähnt auch eine Synchronisation von Veränderungen der Lautstärke und nonverbalem Verhalten. Betonende Handbewegungen werden oft zeitlich mit betonten Silben in Verbindung gebracht und auch das Hochziehen der Augenbrauen wird als mit Rhythmus und Intonation synchronisiert bezeichnet (z.B. bei Santi, 2003). Boyer (2003) beobachtet, dass unterstreichende Handbewegungen zeitgleich mit betonten Silben produziert werden, wohingegen abbildende Gesten am linken Rand lexikalischer Elemente auftreten (vgl. Inhoffen, 1996, die ikonische Gesten unabhängig von der verbalen Sprachproduktion beobachtet, rhythmische unterstreichende Gesten jedoch immer zeitgleich mit der betonten Stelle im Satz und Füllwörter meist ohne begleitende Gestik). Wie Kendon (1980[1977]) stellt Boyer fest, dass auch während Sprechpausen gestikuliert wird, nonverbale Kommunikationsmittel also während eines potentiellen Planungsprozesses verwendet werden. Wallbott (2003a: 274) betont, dass für die Zusammenhänge zwischen prosodischen Strukturen und nonverbalen Mitteln „die Evidenz [...] bisher noch nicht völlig überzeugend scheint“; McClave (1994) spricht gar von einem eigenen, von der verbalen Sprachproduktion unabhängigen Rhythmus sogenannter *Beats* (ursprünglich betonende Taktstockgesten nach Ekman/Friesen, 1969), für die auch Hinweise auf rhythmische Abstimmung zwischen Gesprächsteilnehmern gezeigt werden können.²¹⁸ Die Verwendung dieser Gesten ist nicht von der Beschaffenheit des verbalen Sprachmaterials (Betonungen) abhängig.

²¹⁷ Kendon (1980: 221) verweist auf Birdwhistell (1970), der allerdings nur indirekte Anmerkungen macht.

²¹⁸ Vgl. das Verständnis von Rhythmus im Gespräch nach Auer/Couper-Kuhlen (1994) und Auer et al. (1999).

Was die Entstehung von Sprache anbelangt wird teilweise davon ausgegangen, dass eine Gestensprache als Vorläufer der Lautsprache existierte, aber auch, dass Sprache aus vokalen, handlungsbegleitenden Gesten entstanden ist (s. Nöth, 2002: 302 f.).

In den letzten Jahrzehnten hat nicht zuletzt aufgrund der zunehmenden Globalisierung das Interesse an Untersuchungen zur Körpersprache zugenommen. Im Zuge des wachsenden Interesses an diesem Thema in der Fremdsprachendidaktik erscheinen immer mehr Materialien, die Aspekte wie Gestik in den Fremdsprachenunterricht zu integrieren suchen.²¹⁹ Kulturvergleichende Betrachtungen nonverbaler Mittel sind auch für die Wissenschaft der interkulturellen Kommunikation relevant und werden verstärkt in Seminaren und Trainingskursen zu Interkulturalität behandelt. Besonders Höflichkeit, aber auch die allgemeine Abwicklung von Geschäften werden hier thematisiert (vgl. Rothlauf, 2009).

7.3 Teilbereiche und Klassifizierungen nonverbaler Kommunikation

Unter dem Begriff „nonverbale Kommunikation“ sollen hier die Bereiche Gestik, Kinesik, Mimik, Blickverhalten, taktile Kommunikation, Proxemik, Körperhaltung und Chronemik vorgestellt werden. Anschließend wird auf gängige Unterscheidungen der Arten von Gestik eingegangen.

7.3.1 Teilbereiche nonverbaler Kommunikation

Gestik:

Wie Nöth (2002: 298 f.) erklärt, werden unter dem Begriff nicht nur das „semiotische Ausdruckspotential des menschlichen Körpers mittels der Arme, der Hände und des Kopfes“, sondern in weiterem Sinn z.B. bei Hayes (1966) auch alle anderen Körperbewegungen behandelt. Enger gefasst sollen hier unter „Gestik“ nur die Bewegungen der Hände und Arme verstanden werden, wie sie zu kommunikativen Zwecken, die Rede begleitend oder ersetzend, verwendet werden. Bewegungen der Arme zu anderen Zwecken, die vom Gesagten unabhängig sind, werden hier ausgeschlossen (s. aber Gestenkategorien weiter unten).

²¹⁹ S. Fernández-Conde (2005) für einen Überblick im Bereich des Spanischunterrichtes.

Mimik:

Unter Mimik werden Gesichtsausdrücke verstanden. Diese umfassen Bewegungen einzelner Partien des Gesichtes, wenn z.B. die Augenbrauen hochgezogen oder der Mund unabhängig von der Artikulation bewegt wird. Diese Bewegungen drücken oft Gefühle aus, können aber auch zur Betonung einzelner Satzteile verwendet werden. Die Literatur, die kulturelle Unterschiede bezüglich Gestik und Mimik untersucht, ist umfangreich und wird weiter unten besprochen.

Kinesik:

Kinesik beschäftigt sich mit der Bedeutung von Bewegungen des Körpers. Der Bereich überschneidet sich mit anderen Bereichen nonverbaler Kommunikation. Die Betrachtung der Bewegung der Arme z.B. fällt in den Bereich der Gestik, die Bewegung der Augenbrauen in den der Mimik. Auf die Kinesik wird bei der Vorstellung der Ergebnisse der empirischen Studie im achten Kapitel eingegangen, wenn Veränderungen der Körperhaltung bei Nachrichtensprechern behandelt werden. Dabei werden hauptsächlich Bewegungen des Kopfes, aber auch das Konzept „Bewegung“ im Allgemeinen thematisiert.

Körperhaltung:

Auch die Haltung und Bewegung des Kopfes, des Oberkörpers und weiterer Körperteile kann Informationen z.B. bezüglich der Beteiligung oder des (Des-)Interesses von Sprechern und Hörern liefern und damit als Komponente nonverbaler Kommunikation verstanden werden. Wie Barnlund (1975: 25 f.) indirekt anmerkt, können für Kulturen typische Gangarten unterschieden werden. Daneben können unterschiedliche Körperhaltungen bei Begrüßungen (Aufstehen, Verbeugen, etc.) und Vorträgen (Sitzen, Stehen, Position der Hände, z.B. in den Hosentaschen, usw.) beobachtet und nach Sprechergruppen (z.B. Altersklassen, Geschlecht, aber auch nach Kulturen) unterschieden werden. Informationen, die anhand der Körperhaltung entschlüsselt werden können, schließen das Gemüt, Emotionen und Haltungen bzw. Einstellungen ein.

Taktiler Kommunikation:

Kommunikatives Tasten kann im Körperkontakt zwischen Gesprächsteilnehmern beobachtet werden. Die Art des Kontaktes kann dabei verschiedene Informationen übermitteln, deren Gehalt vermutlich relativ eingeschränkt ist. Das Ausmaß des Körperkontaktes ist stark kulturgebunden. Hall (1966) unterscheidet zwischen *high-contact*-Kulturen (z.B. lateinamerikanische, arabische und südeuropäische Kulturen) und

low-contact-Kulturen (ostasiatische und nordeuropäische Kulturen), in denen unterschiedliche Grade an Körperkontakt in der Öffentlichkeit befürwortet werden. Diese kulturellen Unterschiede spiegeln sich auch in Grußformen wider. Neben Hall untersucht u.a. Barnlund (1975, 1989) Kontakt- und Distanzverhalten in verschiedenen Kulturen und arbeitet Unterschiede zwischen der US-amerikanischen und japanischen Kultur heraus. Seinen Beobachtungen zufolge schränken Japaner für gewöhnlich (nicht nur) körperliche Expressivität stärker ein und messen Berührungen weit weniger Wichtigkeit zu. In diesem Zusammenhang fallen auch Unterschiede zwischen Nachrichtensendungen auf. Zwischen kultur- und geschlechtsspezifischen Unterschieden kann eine Interaktion beobachtet werden: In Italien sei taktile Kommunikation zwischen Männern, in Deutschland zwischen Frauen häufiger.²²⁰

Proxemik:

Der Begriff geht auf Hall zurück und beschäftigt sich mit dem Umgang mit Raum in verschiedenen Kulturen. U.a. wird dabei untersucht, welchen Abstand Menschen typischerweise zueinander einnehmen bzw. mit welchen Distanzen zu anderen Individuen sich Menschen wohl fühlen. Für die US-amerikanische Kultur stellt Hall (1966: 116 ff.) genau trennbare Kategorien von Distanzen auf: intime (0-15cm = nah; 15-45cm = fern), persönliche (45-75cm = nah; 75-120cm = fern), soziale (120-215cm = nah; 215-365cm = fern) und öffentliche (365-760cm = nah; ab 760cm = fern) Distanz.²²¹ Nach Watson (1970) lassen sich kategoriell Kontakt- (z.B. arabische, lateinamerikanische und südeuropäische) von Nichtkontaktkulturen (z.B. asiatische und nordeuropäische) unterscheiden. Luccio (1997: 347) merkt an, dass erwartungsgemäß in Kulturen, in denen eine größere Distanz zwischen Personen gepflegt wird, auch Körperkontakt seltener ist.

Blickverhalten:

Die Art, wie ein Gesprächspartner mit dem Blick adressiert wird und wie dieser erwidert wird, kann verschiedene Informationen übermitteln. Signale werden über den Blickkontakt teilweise unbewusst gesendet oder empfangen (was gewiss auch auf weitere körpersprachliche Kommunikationsmittel zutrifft). Auch die Häufigkeit des Blickkontaktes kann bestimmte Inhalte vermitteln, verstärken oder abschwächen. Relevant ist das Blickverhalten auch, zumindest in westlichen Kulturen, für die Diskursstrukturierung als *Turn-Ende-Signal* (s. Kendon, 1980[1977]). Das Blickverhalten ist stark konventionalisiert

²²⁰ Nach Shuter (1977: 304), zitiert in Ostermeier (1986: 88).

²²¹ Angaben in Zoll und Fuß im Original.

und die Missachtung der Normen kann sanktioniert werden. In interkulturellen Kommunikationssituationen führen Unterschiede im gewohnten Blickverhalten immer wieder zu Missverständnissen und Konfrontationen (s.u.). Für die Untersuchungen der Nachrichtensendungen sind kaum Unterschiede zu erwarten.

Chronemik:

Eine weit verbreitete Unterscheidung von Kulturen nach deren Umgang mit und Auffassung von Zeit ist die Halls (u.a. 1983): Polychrone Kulturen neigen dazu, Zeit als zyklisches Phänomen zu interpretieren und keinen streng geordneten Ablauf von Handlungen zu planen, wie dies für monochrome Kulturen typisch ist. Im Kulturkontrast wurde herausgearbeitet, Personen mit monochronem Zeitverständnis gleichzeitig ablaufende Handlungen und ungenaue Vorausplanung als ungeordnet interpretieren und tendenziell negativ bewerten können. Etwas fragwürdig bleibt die Integration des Zeitverständnisses in den Bereich der (nonverbalen) Kommunikation, spielt jedoch beim Umgang mit Schweigen eine Rolle, das dem kulturellen Umfeld entsprechend verschiedene Reaktionen hervorrufen kann (z.B. Ungeduld bei Konfrontation mit einem ungewohnten Umgang mit Zeit).

Aus semiotischer Sicht sollte darauf hingewiesen werden, dass im Bereich nonverbaler Kommunikationsmittel v.a. in interkulturellen Situationen grundsätzlich die Möglichkeit besteht, dass Zeichen empfangen und dekodiert werden, die vom Sender nicht bewusst gesendet werden, und dass diese Möglichkeit im Vergleich mit Kommunikationsmitteln verbaler Art größer sein mag. Es handelt sich dann um nicht intendierte Kommunikation; im gegenteiligen Fall (gesendete Zeichen werden nicht dekodiert) um nicht gelungene Kommunikation. Bewusst als alleiniges Kommunikationsmittel werden nonverbale Zeichen vermutlich nur in besonderen Situationen, in denen kein (anderer) gemeinsamer Zeichenkode zur Verfügung steht, eingesetzt.

7.3.2 Klassifizierungen von Gesten

Die wohl bekannteste Einteilung von Gesten in Unterklassen ist die von Ekman/Friesen (1969, 1972), in der zwischen Emblemen, Illustratoren, Regulatoren, Affektäußerungen und Körpermanipulatoren unterschieden wird. Embleme bezeichnen dabei Gesten, die einen Begriff darstellen und abbildenden Charakter sowie bei der Sprachproduktion ersetzende Funktion haben können (z.B. Schulterzucken). Illustratoren, die in einige Unterklassen unterschieden werden, betonen Teile lautsprachlicher Äußerungen und haben untermalenden Charakter (z.B. wenn zur Betonung mit gestrecktem Zeigefinger ein Punkt

vor dem Körper in der Luft fixiert wird, eine sogenannte „Taktstockgeste“). Regulatoren regulieren den Diskurs und können strukturierende Funktion im *turn-taking*-Verhalten übernehmen (z.B. abwehrende Handhaltung, wenn man eine Unterbrechung abweist). Als Affektäußerungen werden nicht-verbale Emotionsäußerungen bezeichnet (z.B. ein enttäuschter Gesichtsausdruck). Manipulatoren weisen keinen direkten Bezug zur Sprache auf; es handelt sich um Handlungen, mit denen etwas an einem Objekt oder dem Körper des Sprechers verrichtet wird (z.B. Kratzen). Bei dieser Klassifizierung wird nicht zwischen Gestik und Mimik unterschieden und auch Handlungen, die nicht zu kommunikativen Zwecken eingesetzt werden, finden Berücksichtigung.

Eine weitere bekannte Einteilung ist die nach Freedman et al. (1973), in der zwischen objekt- und körpergerichteten Gesten unterschieden wird. Objektgerichtete Gesten können sprachliche Dominanz aufweisen, wenn sie z.B. als rhythmisierende Gesten im Einklang mit syntaktischen Strukturen produziert werden; sie können andererseits vorrangig motorischen Charakter haben, wenn sie z.B. die sprachlich gelieferte Information abbildend untermalen. Darüber hinaus wird zwischen Gesten unterschieden, die eine ergänzende oder ersetzende Funktion aufweisen. Körpergerichtete Gesten weisen keinen direkten Bezug zur Sprache auf.

In einem Vergleich von Müller (1998: 101 f.) der Klassifizierungen von Efron (1972), Ekman/Friesen (1969), Freedman et al. (1973) und McNeill (1985, 1992) stellt sich heraus, dass alle Autoren zwischen begleitenden und ersetzenden Gesten unterscheiden. Weitere häufige Unterscheidungen sind die zwischen diskursstrukturierenden und gegenstandsbezogenen Gesten sowie zwischen Selbstberührungen und freien Gesten. Müller merkt an, dass keine der Klassifizierungen Gestik objektiv als Kommunikationssystem untersucht, sondern den Gegenstand von spezifischen Fragestellungen ausgehend betrachtet. Zur grundlegenden Unterscheidung bezieht sie sich auf die Formgestalt von Gesten und trennt Selbstberührungen von freien Gesten.

In dieser Arbeit wird zwischen motivierten und opaken Gesten unterschieden. Als motivierte Gesten sollen Gesten bezeichnet werden, die ein Konzept erkennbar, also mit einem transparenten Bezug zum Bezeichneten, abbilden. Sie weisen ikonischen, symbolischen oder indexikalischen Charakter auf und überschneiden sich mit Emblemen (da es sich um konkrete, potentiell sprachersetzende Zeichen handelt) und Illustratoren (da sie deiktisch, piktographisch und strukturgebend sein können; Bsp. s. Kap. 8). Opake Gesten dagegen können nicht von Empfängern entschlüsselt werden, die keinen

gemeinsamen Zeichencode mit dem Sender teilen und keine Hinweise auf den konkreten Kontext haben.

7.4 Kulturelle Unterschiede in nonverbaler Kommunikation

In einigen Arbeiten wird der Frage nachgegangen, inwiefern Gestik und/oder Mimik von Natur oder Kultur determiniert sind. Der Gedanke liegt nahe, dass sie als kommunikative Zeichen ebenso wie verbale Zeichen arbiträren Konventionen unterliegen, die sich in Sprachgemeinschaften etablieren. Einige dieser Zeichen können dabei durchaus transparent und motiviert sein.

7.4.1 Gesichtsausdrücke: Universalien und kulturelle Normierungen

Keltner/Ekman (2000) fassen Ergebnisse früherer Untersuchungen zusammen und führen vier Argumente an, warum bestimmte Gesichtsausdrücke als universell zu bezeichnen sind. In verschiedenen Untersuchungen erkennen Versuchsteilnehmer mit unterschiedlichem kulturellen Hintergrund zumindest sechs von Modellen mimisch dargestellte Emotionen mit einer Übereinstimmung über der Zufallsgrenze, nämlich Wut, Verachtung, Ekel, Angst, Trauer und Überraschung. Außerdem kann nachgewiesen werden, dass sowohl Säuglinge als auch Erwachsene in verschiedenen Kulturen mit ähnlichen Gesichtsausdrücken auf körperliche Einschränkungen (Fixierungen der Arme) reagieren und ähnliche mimische Reaktionen in bestimmten Situationen zeigen (z.B. wenn Wut in Auseinandersetzungen aufkommt). Weiterhin zeigen Versuchsteilnehmer unterschiedlicher Kulturen ähnliche Gesichtsausdrücke, wenn sie selbst gebeten werden, bestimmte Situationen mimisch darzustellen. In Umfragen kann außerdem festgestellt werden, dass Reaktionen von Anhängern unterschiedlicher Kulturen den gleichen Situationen zugeordnet werden, z.B. nervöses Lächeln bei Verlegenheit. Obwohl manche Untersuchungen teilweise methodische Mängel aufweisen und Einflüsse von Konsequenzen der zunehmenden Globalisierung auf die Ergebnisse nicht in jedem Fall ausgeschlossen werden können, darf man davon ausgehen, dass gewisse Gesichtsausdrücke universellen Charakter haben. Dem ist hinzuzufügen, dass nach einer Untersuchung von Eibl-Eibesfeldt blind und taub geborene Kinder die gleichen Muster in ihrer Mimik aufweisen wie sehende und hörende Kinder, wenn Gefühle wie Freude, Trauer und Zorn empfunden werden (vgl. 1999: 679 f.). Vorausgesetzt wird hier, dass diese Emotionen wie bei anderen Kindern empfunden werden. Da diese Kinder die Mimik nicht über den visuellen Kanal oder anhand von Erklärungen gelernt haben können, kann von angeborenem Verhalten ausgegangen werden.

Gleichzeitig belegen Keltner und Ekman (2000) anhand der Ergebnisse von vier Untersuchungen die Existenz von *display rules*. Diese normieren den Gebrauch von Gesichtsausdrücken in Abhängigkeit von kulturellen Konventionen. So wird die Intensität von Gesichtsausdrücken, die Versuchsteilnehmer zu interpretieren haben, von Japanern systematisch anders eingeschätzt als von US-Amerikanern. Weiterhin werden unterschiedliche Rückschlüsse auf Charaktereigenschaften anhand von bestimmten Gesichtsausdrücken, z.B. Lächeln gemacht, das von US-Amerikanern mit Geselligkeit in Verbindung gebracht wird – unter japanischen Versuchsteilnehmern wird diese Assoziation seltener erwähnt. Die Darstellung von Gesichtsausdrücken wird außerdem mit unterschiedlichen Situationen in Verbindung gebracht: Versuchsteilnehmer aus Indien und den USA erwähnen unterschiedliche Gründe, wenn sie gebeten werden, das Lächeln auf dem Gesicht von Modellen zu erklären. Schließlich scheinen auch Unterschiede in der Latenzzeit zu bestehen: US-Amerikanische Kinder reagieren schneller mit Ärger als Japanische. Die überzeugendste und zugleich bekannteste Untersuchung (Ekman, 1973) belegt unterschiedliche Normierungen von Gesichtsausdrücken zwischen einer japanischen und einer US-amerikanischen Testgruppe. Den Versuchsteilnehmern wurde ein Film mit unangenehmen Szenen in zwei Situationen vorgespielt: in Anwesen- bzw. Abwesenheit des Versuchsleiters. Nur unter den japanischen Teilnehmern ist eine deutliche Tendenz zu beobachten, bei Anwesenheit des Versuchsleiters keine negativen Gesichtsausdrücke, teilweise auch positive Gesichtsausdrücke (Lächeln) zu zeigen. Die Studie führte zu der Erkenntnis, dass in einigen ostasiatischen Kulturen der Ausdruck negativer Emotionen zumindest gegenüber Höhergestellten teilweise gesellschaftlich sanktioniert und daher eingeschränkt wird.

Zusammenfassend kann festgehalten werden, dass einige Embleme als Universale angenommen werden können, die Verwendung dieser und anderer nonverbaler Kommunikationsmittel aber kulturell normiert ist. Anschließend sollen qualitative und quantitative Unterschiede in den Bereichen Gestik und Mimik dargestellt werden.

7.4.2 Gestik: Qualitative Unterschiede

Wissenschaftliche Untersuchungen decken Besonderheiten bezüglich der Existenz, Gestalt und Bedeutung einzelner Embleme auf. So zeigt Müller (1998), dass in ihrem Korpus (jeweils zehn Dialoge unter Spaniern bzw. Deutschen von maximal 20 min Länge) Spanier tendenziell mehr diskursive (aufzählende, rhythmische und unterstreichende) Gesten, Deutsche dagegen mehr performative (präsentierende und abwinkende) Gesten verwenden (vgl. Inhoffen 1996: 62, die in Unterhaltungen zwischen Spaniern fast 90% der produzierten Gesten als „gestos rítmicos“ bezeichnet). Außerdem werden andere Informationen gestisch dargestellt: Bei Gesten, die Bewegungsverben begleiten, sind bei den Spaniern Zeichen zu beobachten, die den Weg darstellen, wohingegen von deutschen Sprechern häufiger die Art und Weise der Bewegungsart abgebildet wird.

Eine wertende Beschreibung der Gestenverwendung in Spanien liefert Beinhauer (1930: 204) und grenzt die Verwendung von Gesten, die Ausdrücke begleiten, „wenn es sich um Dinge handelt, die außerhalb jeder sinnlichen Reichweite liegen“ auf „primitive Mensch[en]“ ein. Ein Zusammenhang zwischen der Gestenverwendung und dem Bildungsgrad der Sprecher soll hier nicht thematisiert werden.²²² Weiter unten wird jedoch auf die Sprechsituation eingegangen und es wird untersucht, ob diaphasische Gegebenheiten quantitative Unterschiede in der Gestenverwendung erklären können.

Im Folgenden werden Beispiele von Emblemen vorgestellt, die nur in einem begrenzten Kulturraum verwendet werden.

7.4.2.1 Embleme

Dabei können drei Fälle von qualitativen Unterschieden differenziert werden: Solche, in denen gleiche Inhalte anders ausgedrückt werden; solche, in denen gleiche Zeichen andere Bedeutung aufweisen; und solche, in denen bestimmte nonverbale Zeichen keine Entsprechung in anderen Kulturen finden. Hier werden nur einige wenige Beispiele zur Veranschaulichung angeführt.

²²² In diesem Kontext kann hinzugefügt werden, dass Graham/Argyle (1975: 65 f.) über den Zusammenhang zwischen *Intelligenz* und Verwendung nonverbaler Kommunikationsmittel schreiben: „However, in view of the correlations which have been found between cognitive complexity and the use of gesture (Baxter *et al.*, 1968) and between verbal intelligence and sensitivity to decoding nonverbal cues (Davitz, 1964), it seems unlikely that less intelligent subjects would benefit from the use of gesture.“

7.4.2.1.1 Inhalt, der mittels verschiedener Zeichen ausgedrückt wird

Selbst bei grundlegenden Emblemen wie gestischen Symbolen für „ja“ und „nein“ fallen kulturelle Unterschiede auf. Darwin (1872) sieht den Ursprung für die uns bekannten Bewegungen des Kopfschüttelns und Nickens in den Bewegungen des Säuglings, der sich der Mutterbrust zuwendet oder sich von ihr abneigt. LaBarre (1972[1947]) zählt einige weitere Symbole auf, die stattdessen verwendet werden.²²³ In manchen Ländern, z.B. Bulgarien, werden die Bewegungen für „ja“ und „nein“ umgekehrt verwendet.

Weiterhin kann ein Unterschied beobachtet werden, wie eine Aufforderung, sich dem Sprecher bzw. der das Zeichen sendenden Person anzunähern, ausgedrückt wird. „Komm her“ wird in Deutschland mit nach oben gerichteter Handfläche signalisiert, von der die Finger ausgestreckt sind. Dabei wird mit allen, Zeige- und Mittelfinger oder nur dem Zeigefinger eine heranziehende Bewegung nachgeahmt. In vielen anderen Regionen wird bei dieser Geste die Handfläche nicht nach oben, sondern nach unten gerichtet und alle Finger oder auch die ganze Hand zum Körper hin eingerollt. Urtel (2001[1928]: 38) schreibt zur Verbreitung der beiden Formen, dass die „nördliche Form“ mit nach oben gerichteter Handfläche in Frankreich, Deutschland und Norditalien üblich ist, während die mit nach unten gerichteter Handfläche u.a. in den südlichen Mittelmeerregionen und arabischen Ländern verwendet wird. Nach Meo-Zilio/Mejía (1980) dagegen werden beide Formen dieser Geste in einigen spanischsprachigen Ländern Lateinamerikas verwendet, teilweise mit Bedeutungsunterschied. So wird die mit der Handfläche nach unten gerichtete Form Personen gegenüber verwendet, mit denen die vertraute Anrede gepflegt wird, wohingegen die mit der Handfläche nach oben gerichtete Form, die, zumindest dem Foto auf S. 16 nach zu urteilen, mit nach oben abgewinkeltem Unterarm ausgeführt wird, Personen gegenüber verwendet wird, an die man sich mit respektvolleren Anredeformen wendet. Für Kolumbien wird beschrieben, dass die zweite der beiden Formen von Höhergestellten verwendet wird, um mehrere Personen aufzufordern, sich anzunähern. In Mexiko wiederum weist diese zweite Form eine provokative Komponente auf. Auch die Form, nur einen oder zwei Finger zu bewegen, ist offensichtlich in Lateinamerika bekannt und weist ebenfalls Bedeutungsunterschiede je nach Region auf.

Deiktische Gesten zum Verweis auf den Ort, an dem sich Gegenstände befinden, nach denen gefragt wird, werden ebenfalls in unterschiedlichen Formen verwendet. Sherzer

²²³ Urtel (2001[1928]: 37) erwähnt außerdem für Spanien das Zurückwerfen des Kopfes, das von einem Schnalzlaut begleitet wird, als Zeichen für Ablehnung/Verneinung.

(1973, zu den Kuna-Indianern in Panama) sowie LaBarre (1972[1947], zu den Kiowa in Nordamerika) berichten von Zeigebewegungen, die bei Indianerstämmen mit den Lippen ausgeführt werden (vgl. Enfield, 2001, zu Sprechern aus Laos). Selbst Zeigegesten, die als motiviert und transparent gelten, dürfen nicht als universell bezeichnet werden.

Darüber hinaus soll noch erwähnt werden, dass unterschiedliche Arten, mit den Fingern zu zählen, verwendet werden. Dabei kann mit dem Daumen, Zeigefinger oder kleinen Finger zu zählen begonnen werden. In China (und anderen Ländern) werden für Zahlen ab fünf Gesten verwendet, die für westliche Beobachter vollkommen opak sind. Zu verschiedenen Zählweisen mit den Fingern s. Meo-Zilio/Mejía (1983: 79 ff.).

7.4.2.1.2 Zeichen, die unterschiedliche Bedeutungen ausdrücken

La Barre (1972[1947]) erklärt unterschiedliche Bedeutungen für das kurze Herausstrecken der Zunge. Im südchinesischen Raum bringt dies zum Ausdruck, dass der Sender einen Fauxpas eingesteht; in europäischen Kulturen wird es eher als Beleidigung interpretiert und in der Kultur der Maya muss von einer anderen, unentschlüsselten Bedeutung ausgegangen werden, da das Zeichen oft auf Götterstatuen zu sehen ist.

Wird der Daumen von der zur Faust geballten Hand nach oben abgespreizt, kann das in Brasilien ausdrücken, dass der Sender dem Empfänger viel Glück wünscht (s. Harrison, 1983: 101). In Europa wird dieses Emblem eher dazu verwendet, besonders positive Qualität zu umschreiben.

Für den nicht-gestischen Bereich wird oft darauf hingewiesen, dass der Blickkontakt je nach kulturellem Hintergrund des Empfängers unterschiedlich interpretiert wird. In diesem Zusammenhang wird der direkte Blickkontakt von Personen mit Wurzeln in arabischen Kulturen genannt (vgl. Mehrabian, 1972), der z.B. von US-Amerikanern als aggressiv oder herausfordernd empfunden werden kann.

7.4.2.1.3 Zeichen, die keine Entsprechungen finden

In Spanien und einigen weiteren romanisch geprägten Kulturen wird eine Geste verwendet, die ausdrückt, dass sich an einem Ort besonders viele Menschen befinden. In Spanien werden dabei die Finger einer Hand oder beider Hände nach oben zeigend zusammengeführt und dann abgespreizt und wieder zusammengeführt (s. Gelabert/Martinell Gifre 1990: 23). Im Deutschen Sprachraum ist hierfür kein entsprechendes Emblem verbreitet. Auch das Zeichen des Gehörnten, eines von der Partnerin betrogenen Mannes, drückt einen Inhalt aus, der im deutschen Sprachraum nicht gestisch dargestellt wird. Aus Mexiko sind außerdem eigentümliche gestische Zeichen (z.B. für Bejahung,

wenn der Sprecher während des Essens an der verbalen Sprachproduktion gehindert ist) bekannt.

Anschließend wird auf einen besonderen Unterschied in der Beschaffenheit von Gesten eingegangen: den Gestenraum.

7.4.2.2 Gestenraum

Neben den oben erwähnten Ergebnissen findet Müller (1998) auch Unterschiede in Richtung und Abstand der Gliedmaßen vom Körper während des Gestikulierens. Der Gestenraum der spanischen Sprecher ist größer: Nur sie verwenden auch den Raum über dem Kopf. Auch die Entfernung der Gliedmaßen vom Körper ist bei Spaniern insgesamt weiter als bei Deutschen.

Efron (1972[1941]) untersucht vier kulturelle Untergruppen anhand von 2810 Personen und findet ähnliche Unterschiede zwischen „ostjüdischen“ und italienischen nicht-assimilierten Einwanderern in New York. Der Gestenraum der ostjüdischen Sprecher ist deutlich kleiner und auf den Bereich vor und neben dem Oberkörper eingeschränkt, wohingegen bei italienischen Sprechern weiter ausgreifende Bewegungen auch über dem Kopf zu beobachten sind.

Nachdem Gesellschaften ihrem proxemischen Verhalten entsprechend unterschieden wurden, liegt auf der Hand, dass parallele Unterschiede in der nonverbalen Kommunikation zu beobachten sind und der vom Individuum in seiner Gestik in Anspruch genommene Raum unterschiedlich groß ist.

7.4.3 Gestik und Mimik: Quantitative Unterschiede und Erklärungsansätze

Zeichensysteme nonverbaler Mittel unterscheiden sich nicht nur in der Beschaffenheit und Bedeutung der verwendeten Zeichen – auch bei der Häufigkeit der Verwendung sind Unterschiede zu beobachten. Dabei spielen viele Faktoren eine Rolle, von individuellen Gewohnheiten über diaphasische Gegebenheiten bis zu kulturellen Konventionen. Das Alter scheint hierbei ebenso von Bedeutung zu sein: Mit zunehmendem Alter nimmt die Zahl ikonischer Gesten nach Cohen/Borsoi (1996) ab. Die Zahl rhythmisierender Gesten dagegen nimmt zu, so dass eine Abnahme der Bedeutung visueller Zeichen vermutet werden kann (s. Wallbott 2003a: 274). Prinzipiell kann davon ausgegangen werden, dass die im zweiten Kapitel erwähnten Faktoren auch auf die Verwendung nonverbaler

Kommunikationsmittel einen Einfluss haben. Im Folgenden wird der kulturelle Aspekt näher betrachtet.

Japaner wurden im Vergleich mit US-Amerikanern als weniger expressiv bezeichnet, was die motorische Aktivität beim Sprechen und den Ausdruck von Emotionen anbelangt.²²⁴ Harrison (1983: 92 f.) betont wiederum, dass Brasilianer häufiger gestikulieren als US-Amerikaner. Eine Tendenz zur häufigen Benutzung nonverbaler Mittel wurde für südamerikanische, osteuropäische, arabische, schwarzamerikanische und Kulturen aus dem Mittelmeerraum postuliert und teilweise durch empirische Arbeiten belegt (vgl. Matsumoto et al., 2002). Nordeuropäische und asiatische Kulturen dagegen gelten generell als Kulturen, in denen weniger gestikuliert wird.²²⁵ Wandruszka (2001[1954]: 105) stellt „lebendigen“ Romanen „steife“ Nordländer gegenüber. Im Allgemeinen wird also für Kulturen, die als *high-contact*-Kulturen gelten, von einer häufigeren Verwendung nonverbaler Zeichen ausgegangen. Für diese Kulturen kann eine stärkere Bedeutung der nicht-akustischen Kanäle vermutet werden (vgl. Barnlund, 1975). Konkreter verwenden Graham/Argyle (1975: 58) die Begriffe „*high gesture culture*“ und „*low gesture culture*“ und erklären damit Unterschiede in Quantität und Qualität der Gesten von Italienern und Engländern. Dabei verweisen sie auf Schefflen (1972), der betont, dass die Verwendung von Gesten während der Erziehung US-amerikanischer Kinder britischer Herkunft eingeschränkt und sogar unterbunden wird. In ihrer Untersuchung finden Graham und Argyle heraus, dass die Beschreibung von konkreten und abstrakten Formen unter Zuhilfenahme von Gesten besser verstanden wird, als wenn Gesten nicht verwendet werden. Dies trifft sowohl auf englische als auch auf italienische Versuchsteilnehmer zu, wobei unter Letzteren eine wesentlich deutlichere Verbesserung unter der Kondition „mit Verwendung von Gesten“ eintritt und die beschriebenen Formen von Hörern genauer nachgezeichnet werden können (vgl. Walker/Nazmi, 1979, zur Erinnerung abstrakter Formen, die mit und ohne begleitende Gesten beschrieben werden und von italienischen Versuchsteilnehmern besser als von Versuchsteilnehmern mit englischem Hintergrund erinnert werden, wenn Gesten die Beschreibungen begleiten). Da die verbale Flüssigkeit in beiden Konditionen vergleichbar ist, schließen die Autoren aus den Ergebnissen, dass für italienische Versuchsteilnehmer nonverbale Kommunikationsmittel von größerer Bedeutung sind als für Englische. Es muss davon ausgegangen werden, dass die Verarbeitung

²²⁴ Porter/Samovar (1998: 464), s. auch Barnlund (1975: 105).

²²⁵ Vgl. Kommentare in Andersen (2002), Kendon (1983), Inhoffen (1996), Wierzbicka (1999, 2003).

von Gestik, genauer: die Dekodierungsleistung in Abhängigkeit gestischer Zeichen, je nach Sprachgemeinschaft variiert.

Gleichzeitig liegen Hinweise darauf vor, dass keine quantitativen Unterschiede bezüglich der Verwendung von Gestik zwischen Gruppen bestehen, zwischen denen solche aufgrund ihrer Zurechnung zu unterschiedlichen *gesture*-Kulturen vermutet wurden; zumindest gilt das für vergleichbare Sprechsituationen: „Spanische Sprecher gestikulieren nicht mehr als deutsche, womit das Stereotyp des 'mit Händen und Füßen redenden' Südeuropäers zumindest für redegleitende Gesten bei spontanen Schilderungen von Bewegungsereignissen widerlegt [sic] wäre“, fasst Müller (1998: 231) ihre Ergebnisse zusammen. Sicherlich spielt selektive Wahrnehmung eine Rolle, wenn Spanier als „mit Händen und Füßen redend“ empfunden werden. Es ist vorstellbar, dass die Dekodierung vertrauter nonverbaler Mittel so automatisiert ist, dass nur unbekannte Gesten und Gesichtsausdrücke bewusst wahrgenommen werden und dadurch der falsche Eindruck entsteht, Gesprächspartner mit einem anderen kulturellen Hintergrund verwenden *mehr* Gesten, obwohl sie hauptsächlich *andere* Gesten verwenden. Dass dieser Eindruck offensichtlich nur für Sprecher in sogenannten *low gesture*-Kulturen entsteht (Schilderungen, z.B. von Südeuropäern, die Nordeuropäer als „mit Händen und Füßen redend“ bezeichnen, sind nicht bekannt), kann damit erklärt werden, dass diese über ein begrenztes nonverbales Zeichenrepertoire verfügen und bei Sprechern, die sich eines umfangreicheren nonverbalen Zeichenrepertoires bedienen, neue, unbekannte Zeichen beobachten. Umgekehrt müssten die Sprecher aus *high-gesture*-Kulturen mit Zeichen, die in *low-gesture*-Kulturen verwendet werden, weitgehend vertraut sein und daher keine auffallende Verwendung dieser Zeichen wahrnehmen. Auch die Verwendung eines größeren Gestenraumes seitens spanischer Sprecher kann bei deutschen Gesprächsteilnehmern die Wahrnehmung intensiverer Gestik auslösen, die dann als häufigere Gestik missinterpretiert wird. Dennoch scheint die Verwendung nonverbaler Zeichen kulturell normiert und wird nicht in allen Sprachgemeinschaften gleichermaßen gebilligt. Für die Wahrnehmung nonverbaler Zeichen spielen Stereotype und Vorurteile sicherlich eine bedeutende Rolle.

Wierzbicka sucht Erklärungen für derartige Unterschiede und beschreibt unter Verwendung der „Natural Semantic Metalanguage“²²⁶, welche Einstellungen und

²²⁶ Bei der NSM handelt es sich um eine Reihe von universellen Begriffen, die semantischen Primitiven entsprechen. Mit ihr werden Konzepte objektiv beschrieben um einen terminologischen Ethnozentrismus zu vermeiden. Sie werden für die Formulierung von sog. *cultural scripts* verwendet, die wiederum erklären sollen, aufgrund welcher Haltungen und (impliziter) Annahmen sich gewisse Diskurskonventionen etablieren.

Annahmen in kulturellen Systemen zugrunde liegen, die bestimmte Verhaltensmuster auslösen. Für die polnische Kultur, die als „expressiv“ bezeichnet wird, formuliert sie (1999: 271) das *cultural script*:

(7.1)

[everybody knows:]
often when a person is saying something
other people can know what this person feels
because they can see this person's face
[people think: this is good]

In Bezug auf Proxemik und taktile Kommunikation geht sie auf S. 253 f. von diesem *cultural script* aus:

(7.2)

[everybody knows:]
often when people feel something good towards someone else
they do something with their body at the same time
they want to do it
when they do it parts of their body touch parts of the other person's body
because of this other people can know what these people feel
[people think: this is good]

Diesen Skripten liegen die folgenden Annahmen zugrunde:

(7.3)

In different societies, and different communities, people speak differently.
These differences in ways of speaking are profound and systematic.
These differences reflect different cultural values, or at least different hierarchies of values.
Different ways of speaking, different communicative styles, can be explained and made sense of, in terms of independently established different cultural values and cultural priorities.

(nach Wierzbicka 2003: 69)

Weiterhin erklärt Wierzbicka (1999: 266), dass für Personen mit anglo-amerikanischem kulturellem Hintergrund das Konzept persönlicher Autonomie von Bedeutung ist.

(7.4)

[people think:]
when I do something I want to know:
"I do it because I want to do it
(not because I feel something)"

Dieses grundlegende Konzept begünstigt, dass eine weitere Haltung in diesem Kulturkreis von Bedeutung ist:

(7.5)

[people think:]

it is good if other people know what I feel when I want them to know it

it is good to say what I feel when I want other people to know it

Diese kulturellen Grundeinstellungen können erklären, wie es zu unterschiedlichen Graden an Expressivität an der Oberfläche kommunikativen Handelns kommt: einerseits „expressive“ Konventionen in der polnischen Kultur, in der der (non-) verbale Ausdruck von Emotionen scheinbar als Normalfall gilt, andererseits „reserviertes“ Verhalten in den angloamerikanischen Kulturkreisen, in denen die Wahrung der eigenen persönlichen Autonomie und der Anderer als vorrangig erachtet wird.²²⁷ Man kann sich kulturelle Skripte als *constraints* in einem kulturell orientierten optimalitätstheoretischen Rahmen vorstellen.

In Matsumoto et al. (2008) wird untersucht, inwiefern die kulturelle Dimension Individualismus-Kollektivismus einen Einfluss auf die allgemeine Expressivität und spezifische, die Expressivität regulierende Normen hat.²²⁸ In der Untersuchung werden mehr als 5000 Studenten in 32 Ländern mittels Fragebögen dazu befragt, wie sie sich in verschiedenen situativen Rahmen verhalten sollten, wenn sie bestimmte Emotionen empfinden. Die Situationen werden nach Gesprächspartner der Eigen- bzw. Fremdgruppe (z.B. Familienangehörige, Hochschullehrer) und Gesprächssituation (alleine, in der Cafeteria) unterschieden. Es stellt sich heraus, dass der Grad an Individualismus (Werte aus Hofstede, 2001) positiv mit der allgemeinen Expressivität korreliert. Innerhalb der Eigengruppe werden in individualistisch geprägten Kulturen allgemein mehr Emotionen gezeigt als in kollektivistisch geprägten. Gegenüber der Fremdgruppe werden in Kulturen mit hohen IDV-Werten weniger negative Emotionen gezeigt, dafür mehr positive (Freude und Überraschung). Außerdem unterscheiden individualistische Kulturen stärker zwischen Eigen- und Fremdgruppe. Die Ergebnisse stehen teilweise im Widerspruch zu den Überlegungen Wierzbickas, nach denen in der polnischen Kultur Expressivität stärker gefördert wird als in der US-amerikanischen. Nach Hofstede (2001: 500-502) weist die

²²⁷ Es sollte vorsichtshalber nicht davon ausgegangen werden, dass die beschriebenen Haltungen uneingeschränkt für ein als angloamerikanischer Kulturraum bezeichnetes Gebiet anzunehmen sind. Dass innerhalb eines derartigen Gebietes ähnliche Ansichten in vielen Kontexten geteilt werden, ist aber nicht von der Hand zu weisen.

²²⁸ Matsumoto et al. (2008: 59) fassen Definitionen von IDV wie folgt zusammen: „[I]ndividualistic cultures foster the development of independent construals of self [...], favor personal goals over in-group goals [...], encourage rationality and interpersonal exchange [...], and place more importance on attitudes as relatively important determinants of behavior.“

US-amerikanische Kultur einen höheren Grad an Individualismus auf als die polnische (91 = Rang 1/54 > 60 = Rang 18/54).

In der Untersuchung von Matsumoto et al. (2008: 63) wird nicht eindeutig zwischen verbalen und nonverbalen Verhaltensmustern unterschieden. So werden z.B. „express it as you feel it“ und „show less than you feel it“ als Antwortmöglichkeiten angegeben, wobei in den Fragebögen nicht weiter spezifiziert wird, ob unter „express“ nur verbales, verbales und nonverbales oder unter Umständen auch nur nonverbales Verhalten zu verstehen ist. Des Weiteren ist nicht gegeben, dass in Kulturen, in denen starke Regulierungsnormen bezüglich des Ausdrucks bestimmter Emotionen gepflegt werden, offen Antworten wie „hide your feelings by smiling“ gegeben werden.

Dass zwischen Individualismus/Kollektivismus und dem Ausdruck von Emotionen ein Zusammenhang besteht, soll nicht in Frage gestellt werden. Sicherlich kann sich Expressivität auf ganz unterschiedliche Arten äußern, so dass Verallgemeinerungen für die Verwendung nonverbaler Kommunikationsmittel kaum haltbar sind.

Da expressive Gesichtsausdrücke von Beobachtern unterschiedlicher Kulturen als unterschiedlich intensiv interpretiert werden (Ekman et al., 1987), kann nicht ausgeschlossen werden, dass Emotionen in verschiedenen Kulturen unterschiedlich stark wahrgenommen werden. Uchida et al. (2009) liefern Hinweise darauf, dass in der japanischen Kultur Emotionen als zwischenmenschliches Phänomen verstanden werden, wohingegen unter US-Amerikanern Gefühle als im Individuum begründet gelten.²²⁹ Damit muss als fragwürdig gelten, ob Emotionen, sofern sie auf gleiche oder zumindest ähnliche Weise nonverbal ausgedrückt werden, auch unter gleichen/ähnlichen Umständen empfunden werden. Derartiges über die als universell geltenden Emotionen (s.o.) hinaus zu postulieren scheint unrealistisch, da Definitionen von Gefühlen, die über einfache körperliche Reaktionen hinausgehen, teilweise innerhalb von Kulturen schon schwierig sind.

²²⁹ In der Untersuchung werden Interviews mit Sportwettkämpfern ausgewertet. Japanische Sportler verwenden mehr emotionale Begriffe, wenn zwischenmenschliche Beziehungen (zum Trainer, dem Team, etc.) thematisiert werden, wohingegen US-amerikanische Sportler bei Fokussierung auf die eigene Person emotionale Begriffe verwenden. Das Ergebnis zeigt auch, wie Vorurteile entstehen können, wenn in Kontexten, in denen in der eigenen Kultur Gefühle gezeigt werden, von Individuen mit einem anderen kulturellen Hintergrund vergleichsweise emotionslose Reaktionen gezeigt werden, da in der anderen Kultur in einer entsprechenden Situation keine emotionale Reaktion üblich ist.

7.5 Zusammenfassung

Es konnte gezeigt werden, dass Gesellschaften ungleiche Verwendungen körpersprachlicher Zeichen pflegen. Es bestehen Unterschiede in der Qualität (sowohl bei arbiträr konventionalisierten als auch bei motivierten, transparenten Zeichen), der Quantität (Häufigkeit der Verwendung nonverbaler Zeichen), der Bedeutung einzelner Zeichen und der Anwendung (dem Kontext) nonverbaler Zeichen. Obwohl eine Reihe von Emotionen und ihr Ausdruck als universell gelten, ist gleichzeitig bekannt, dass bestimmte Emotionen und Einstellungen durch kulturelle Konventionen geprägt sind und Unterschiede in der nonverbalen Kommunikation erklären können. So kann gezeigt werden, dass Informationen über den emotionalen Zustand von Sprechern in manchen Kulturen eher mitgeteilt werden als in anderen, wobei eine Interaktion mit dem situativen Kontext besteht. Innerhalb von Gesellschaften wird zwischen der Eigen- und der Fremdgruppe unterschieden; die entsprechenden Kommunikationssituationen erfordern die Anwendung unterschiedlicher Normen. Die für diese Situationen gültigen Regeln hängen außerdem mit dem Grad an Individualismus zusammen. Es muss auch hier betont werden, dass die Wahrnehmung nonverbaler Zeichen im interkulturellen Kontext von Stereotypen und Vorurteilen beeinflusst wird.

Nachdem die Grundlagen der kulturvergleichenden Betrachtung nonverbaler Zeichen dargelegt wurden, wird im folgenden Kapitel eine empirische Studie zur Qualität und Quantität nonverbaler Zeichen in den untersuchten Nachrichtensendungen beschrieben und ihre Ergebnisse in den Gesamtzusammenhang eingeordnet.

8 Empirische Untersuchung zur nonverbalen Kommunikation

8.1 Einleitung

Die Verwendung von gestischen, mimischen und anderen körpersprachlichen Zeichen ist ein Bestandteil sprachlicher Kommunikation. Wie Manusov/Jaworski (2006) betonen, spielen Medien eine bedeutende Rolle bei der Sozialisierung. Das Fernsehen hat einen nachweislichen Einfluss darauf, wie Kinder lernen, Emotionen auszudrücken. Die Bedeutung der Verwendung nonverbaler Kommunikationsmittel in den Medien für die Tradition gesellschaftlicher Umgangsformen und Kommunikationsstile darf nicht unterschätzt werden. Nachdem festgestellt wurde, dass Sprachen in einer standardisierten Sprechsituation unterschiedlich schnell gesprochen werden, soll nun untersucht werden, welche weiteren relevanten sprachlichen Unterschiede auffallen. Dabei wird für die Untersuchung nonverbaler Kommunikationsmittel nicht die Geschwindigkeit, sondern die Häufigkeit, die auch als Dichte aufgefasst werden kann, betrachtet. Mit höherer Anzahl von nonverbalen Zeichen pro Zeiteinheit steigt die Dichte der gesendeten Informationen, die als Aspekt des Sprechtempos verstanden werden kann. Die mittels Gestik gesendete Information kann dabei inhaltlich mit der verbal gesendeten gleich sein, diese modifizieren oder ergänzen; für die hier untersuchte Sprechsituation ist nicht davon auszugehen, dass sie ihr widerspricht.

Fünf der sechs im vierten Kapitel untersuchten Sprachen wurden auf die Verwendung nonverbaler Mittel in Nachrichtensendungen untersucht. Der Untersuchungsgegenstand umfasst Gestik, Mimik und Kinesik. Dabei ist zu beachten, dass es sich um eine besondere Sprechsituation handelt, bei der keine direkte Interaktion mit dem Empfänger möglich ist. Nachrichtensendungen können aber als Kommunikationssituation von Angesicht zu Angesicht und als Kommunikation zwischen Moderator und Zuschauer konzipiert sein (vgl. Scollon, 1998). Dass der monologische Rahmen durch eine direkte Bezugnahme des Sprechers auf den Zuschauer durchaus gesprengt werden kann, wird weiter unten (in 8.4.2.2) gezeigt. Folgend werden die Methode und das untersuchte Material vorgestellt und anschließend quantitative und qualitative Unterschiede zwischen den Sprachen anhand von Beispielen vor Augen geführt.

8.2 Methode

Für die quantitative Erfassung nonverbaler Mittel wurden in allen Nachrichtensendungen Gesten, mimische Ausdrücke und Veränderungen der Position des Kopfes bzw. des Oberkörpers gezählt und die durchschnittliche Anzahl pro Minute berechnet.²³⁰ Wie im ersten Teil der Arbeit wurden arithmetische Mittelwerte für Sprachen auf Grundlage der globalen Durchschnittswerte der einzelnen Sprecher berechnet. Für die Auszählung der Gesten ergibt sich dabei das Problem, dass motorische Sequenzen teilweise nur schwer in trennbare Einheiten zu unterteilen sind. Eine mehrfach beobachtete strukturierende Geste, die eine inhaltliche Gegenüberstellung begleitet, besteht aus zwei einander entgegengesetzten Seitwärtsbewegungen der Hände oder Unterarme bzw. des Kopfes: Die Begriffe „einerseits..., andererseits...“ oder „zwar..., aber...“ werden mithilfe dieser Bewegungen physisch dargestellt voneinander getrennt und einander gegenübergestellt. Diese Bewegungen bilden also eine gliedernde Struktur motorisch ab, wobei auch beobachtet werden kann, dass nur der erste Bestandteil ausgeführt wird. Die zweite Komponente kann dann durch eine Kopfbewegung oder ein betonendes Heben der Augenbrauen ersetzt werden oder ganz entfallen. Daher werden, obwohl inhaltlich eine Einheit vorliegt, beide Bewegungen in der Auszählung berücksichtigt.

Weiterhin werden rhythmische „Schläge“ mit der Hand in der Luft produziert, die betonte, aber auch unbetonte Silben begleiten können und eine in sich geschlossene Sequenz darzustellen scheinen. Auch hier wird jeder einzelne dieser Schläge berücksichtigt, da der hierdurch hervorgerufene Eindruck der Gliederung beim Zuschauer durch längere Sequenzen verstärkt wird. Das betonende Schlagen des rhythmischen Taktes unterstreicht die Wichtigkeit des Abschnittes stärker, je länger es andauert, womit die qualitative Wichtigkeit quantitativ dargestellt wird und in der Auszählung berücksichtigt werden muss.

In Bezug auf Mimik kann beobachtet werden, dass grundsätzlich zwei mimische Ausdrücke systematisch produziert werden: das Hochziehen einer oder beider Augenbrauen und ein Stirnrunzeln, das von zusammengezogenen Augenbrauen ausgelöst werden kann. Hochziehen und anschließenden Absenken der Augenbrauen wurden als ein Gesichtsausdruck gezählt, da die Bewegungen nicht unabhängig voneinander produziert

²³⁰ Objektbezogene außersprachliche Handlungen, die bei chinesischen und deutschen Moderatoren beobachtet werden können, wenn Blätter beiseite gelegt werden, wurden dabei nicht berücksichtigt.

werden. Auffällige Bewegungen des Mundes wurden nicht als Gesichtsausdrücke gewertet, da sie von der Artikulation abhängige Begleiterscheinungen von Akzent auftreten.

Außerdem wurden Bewegungen des Kopfes und/oder Oberkörpers (Neigungen und Drehungen in der Horizontalen und Vertikalen) gezählt. Die separat erfasste Anzahl der genannten nonverbalen Elemente wurde durch die SZ geteilt.

Es muss eingeräumt werden, dass die qualitative Beschreibung des Materials teilweise auf den Beobachtungen des Autors beruht und in manchen Fällen noch keine statistischen Analysen durchgeführt wurden.

8.3 Material

Tabelle 8.1 gibt einen Überblick über die verwendeten Nachrichtensendungen. Da der „BBC News Bulletin“ nur als Audiodatei gesendet wird, wird das Englische in dieser Untersuchung nicht berücksichtigt. Aufgrund der Ergebnisse des sechsten Kapitels muss bedacht werden, dass die Stilebene einen Einfluss auf die sprachliche Gestaltung in den Nachrichtensendungen haben kann. Es wurden daher zusätzliche Nachrichtensendungen ausgewertet, um mehrere Registerebenen miteinander vergleichen zu können.

Sowohl die Anzahl der Sprecher als auch die Gesamtlänge der Aufnahmen pro Sprache fallen unterschiedlich aus, was auf den Ablauf der Sendungen zurückzuführen ist. In chinesischen Nachrichtensendungen z.B. sind die Sprecher teilweise nur sehr kurz zu sehen, weshalb viele Nachrichtensendungen über einen längeren Zeitraum untersucht wurden. Die vielen verschiedenen Sprecher konnten teilweise nur in kurzen Ausschnitten untersucht werden. Auch der Bildausschnitt (nur Kopf, Kopf und Oberkörper oder ganzer Körper im Bild) schränkt das verwendbare Material ein. Im Fall des BP wurden einige Sendungen nur auf Mimik und Bewegungen des Kopfes hin untersucht, da nur das Gesicht der Moderatoren zu sehen war, weshalb hier das umfangreichste Material vorliegt. Insgesamt wurden 81 Sprecher berücksichtigt.

Zur Unterscheidung von vier verschiedenen Registerebenen wurde hauptsächlich die Sendezeit als Kriterium herangezogen: Die Hauptsendezeit (in Deutschland z.B. 20 Uhr) wurde als formellster Kontext festgelegt. Weitere Programme, die zu anderen Sendezeiten ausgestrahlt werden (z.B. die „Tagesthemen“ oder das „Nachtmagazin“) unterscheiden sich von diesen Sendungen bezüglich Umgangsformen, sprachlicher Mittel und teilweise auch inhaltlich, was das Kriterium der Sendezeit untermauert. Sprachlich weichen Sendungen, die nachts oder am frühen Morgen gesendet werden, durch weniger formelle

Ausdrucksformen von den formelleren Sendungen zur Hauptsendezeit ab, die sowohl im Satzbau als auch in der Wortwahl auffallen. Beispiele hierzu sind eliptische Sätze wie „... und das sind wir“ am Beginn der Sendung (Bsp. aus dem „Nachtmagazin“) und die Verwendung von Modalpartikeln wie in „Falls Sie das auch schon immer einmal wissen wollten“ (ebenfalls aus dem Nachtmagazin, s.u.). Auch was prosodische Eigenschaften anbelangt können Unterschiede festgestellt werden (z.B. häufigere rhetorische Pausen im Deutschen). Unterschiedliche Umgangsformen können anhand der Begrüßung und Verabschiedung beobachtet werden (obiges Beispiel vs. „Guten Abend, meine Damen und Herren“ zu Beginn der „Tagesschau“ um 20 Uhr). Nicht in allen Ländern sind Unterschiede zwischen den Sendungen zu beobachten, die die vier in Tabelle 8.1 angegebenen Registerebenen abdecken.

Tabelle 8.1: Auf die Verwendung nonverbaler Kommunikationsmittel hin untersuchte Nachrichtensendungen

Herkunftsland	Sender	Sendung	Register	Sprecheranzahl	Umfang
Brasilien	Globo	Globo Notícia para o G1	formell		3622 s
		Jornal Nacional	formell		665 s
		Jornal Hoje	semiformell		85 s
		Jornal das Dez	semiformell		143 s
		Jornal da Globo	semiformell		544 s
		Bom Dia Brasil	semiinformell		425 s
				4 w, 7 m	=87 min
China	CCTV	新闻联播	formell		611 s
		朝闻天下	semiformell		905 s
		新闻 30 分	semiformell		500 s
		晚间新闻	semiformell		319 s
		东方时空	semiinformell		221 s
		新闻 1+1	informell		449 s
				16 w, 16 m	=50 min
Deutschland	ARD	Tagesschau (20 Uhr)	formell		2090 s
		Tagesschau (Morgenausgabe)	semiformell		2203 s
		Tagesschau in 100 Sekunden	semiinformell		28 s
		Tagesthemen	semiformell		276 s
		Nachtmagazin	informell		556 s
				6 w, 8 m	=56 min
Portugal	RTP1	Telejornal	formell		1664 s
		Jornal da Tarde	semiformell		1304 s
		Bom Dia Portugal	semiformell		1551 s
		Portugal em Direto	semiinformell		293 s
		30 Minutos	informell		161 s
				6 w, 6 m	=83 min
Spanien	TVE1	Telediario	formell		2425 s
		La noche en 24 horas	informell		445 s
		España en 24 horas	informell		68 s
	La 2	La 2 noticias	semiformell		1029 s
				9 w, 3 m	=67 min

8.4 Ergebnisse

8.4.1 Quantitative Untersuchung

Im Folgenden werden Kastengrafiken präsentiert, um den Einfluss der Faktoren Sprache,²³¹ Geschlecht und Register auf die Verwendung nonverbaler Mittel wie oben definiert optisch darzustellen.

8.4.1.1 Faktor Sprache

Die Abbildungen 8.1-8.3²³² zeigen eine bildliche Darstellung der durchschnittlichen Häufigkeiten von Gesten, mimischen Ausdrücken und Veränderungen der Position des Kopfes und/oder Oberkörpers.

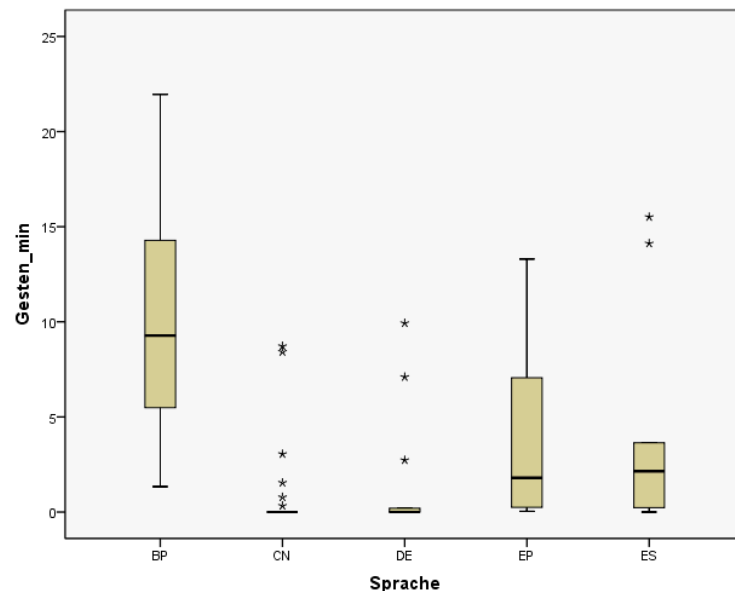


Abbildung 8.1: Häufigkeit der Gesten/min in den untersuchten Sprachen

²³¹ Ebenso könnte der Faktor als „Kultur“ oder „kultureller Hintergrund“ bezeichnet werden. In Anlehnung an die Untersuchung in Kap. 4 werden hier der Begriff „Sprache“ und die bisher gebrauchten Abkürzungen für die einzelnen Sprachen verwendet.

²³² In den Abbildungen sind die in Kap. 4 beschriebenen Informationen enthalten.

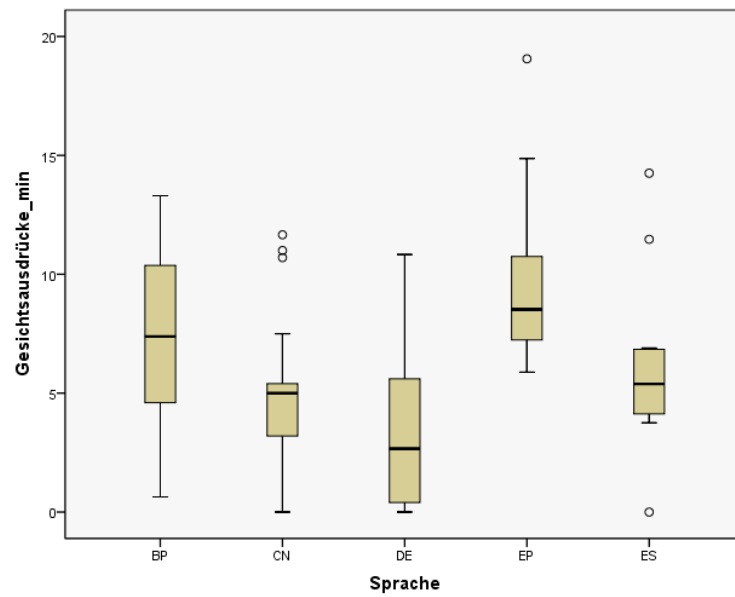


Abbildung 8.2: Häufigkeit der Gesichtsausdrücke/min in den untersuchten Sprachen

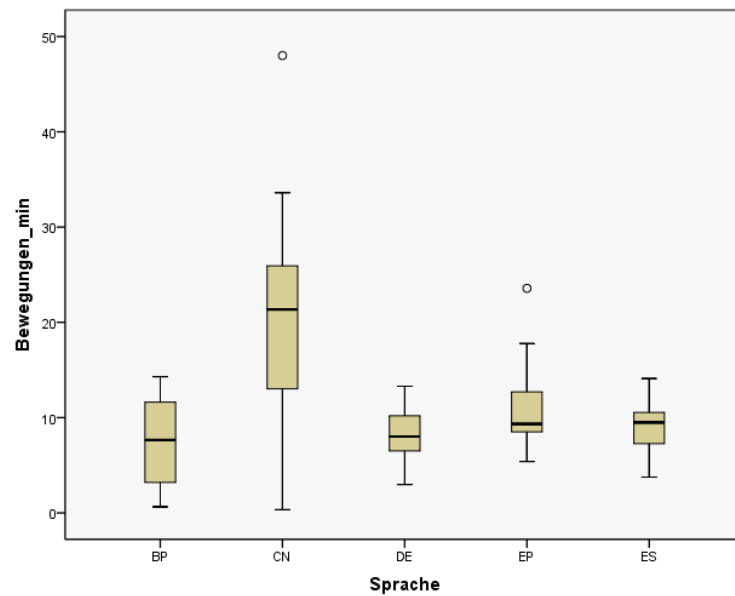


Abbildung 8.3: Häufigkeit der Bewegungen des Oberkörpers und/oder Kopfes/min in den untersuchten Sprachen

Die arithmetischen Mittelwerte sind Tabelle 8.2 zu entnehmen:

Tabelle 8.2: Arithmetische Mittelwerte und Standardabweichungen für produzierte Gesten, mimische Ausdrücke und Veränderungen der Position des Kopfes und/oder Oberkörpers/min in den untersuchten Sprachen

Sprache	Gesten/min		Gesichtsausdrücke/min		Bewegungen/min	
	Mittel	SD	Mittel	SD	Mittel	SD
BP	10,17	6,68	7,33	4,17	7,49	4,86
CN	0,70	2,15	4,44	3,41	19,73	8,44
DE	1,43	3,14	3,44	3,54	8,35	2,91
EP	2,88	2,89	9,59	3,60	10,50	4,15
ES	3,78	5,35	6,18	3,85	9,07	2,92

Für alle drei Variablen sind Unterschiede statistisch signifikant (Gestik: $F_{(4,75)}=13,093$; $p<0,001$; Mimik: $F_{(4,58)}=5,697$; $p=0,001$; Bewegungen: $F_{(4,58)}=12,725$; $p<0,001$). Sehr deutlich unterscheiden sich die brasilianischen Nachrichtensendungen von allen anderen in Bezug auf die Gestenhäufigkeit. Die homogenen Untergruppen lauten:

$$\underline{\text{CN} < \text{DE} < \text{EP} < \text{ES}} < \overline{\text{BP}}$$

Die Darstellung der Häufigkeit mimischer Ausdrücke ergibt eine Überschneidung zweier Untergruppen:

$$\underline{\text{DE} < \text{CN} < \text{ES} < \text{BP}} < \overline{\text{EP}}$$

Anhand der Häufigkeit von Kopfbewegungen lassen sich chinesische Nachrichtensendungen deutlich von den anderen unterscheiden. Die Untergruppen sind:

$$\underline{\text{BP} < \text{DE} < \text{ES} < \text{EP}} < \overline{\text{CN}}$$

8.4.1.2 Faktor Geschlecht

Was geschlechtsspezifische Unterschiede anbelangt ergibt sich kein einheitliches Bild. In deutschen, portugiesischen und spanischen Nachrichtensendungen gestikulieren Männer häufiger als Frauen, mimische Ausdrücke sind im EP häufiger als bei Frauen und die Anzahl der Bewegungen liegt in brasilianischen, chinesischen und spanischen Sendungen bei Frauen höher (s. Tab. 8.3). Keiner der zu beobachtenden geschlechtsspezifischen Unterschiede ist statistisch signifikant, weshalb auf die Darstellung in entsprechenden Kastengrafiken verzichtet wird. Zu weiteren Beobachtungen bezüglich Unterschieden zwischen den Geschlechtern siehe Unterkapitel 8.4.2.5.

Tabelle 8.3: Unterschiede zwischen den Geschlechtern: produzierte Gesten, mimische Ausdrücke und Veränderungen der Position des Kopfes und/oder Oberkörpers/min

Sprache	Geschlecht	Gesten/min		Gesichtsausdrücke/min		Bewegungen/min	
		Mittel	SD	Mittel	SD	Mittel	SD
BP	W	12,48	8,73	8,45	3,38	9,88	5,96
	M	8,85	5,55	6,66	4,81	6,05	4,08
CN	W	0,77	2,13	6,33	3,73	25,36	10,76
	M	0,76	2,40	4,28	2,63	16,31	10,51
DE	W	1,18	2,90	3,98	3,63	7,06	2,41
	M	1,61	3,49	3,03	3,67	9,32	3,01
EP	W	1,53	2,74	8,90	2,46	10,42	4,24
	M	5,13	4,56	10,59	4,74	12,09	6,04
ES	W	2,92	4,46	6,43	4,57	9,53	3,20
	M	6,38	8,05	5,51	0,56	7,87	1,92

8.4.1.3 Faktor Register

Für die Stilebene kann in allen außer den portugiesischen Nachrichtensendungen ein Einfluss auf die Verwendung von Gestik festgestellt werden (s. Korrelationsmatrix in Tab. 8.4 und Abb. 8.4 zur bildlichen Veranschaulichung). Signifikante Unterschiede zwischen den Registerebenen können mit einem einfaktoriellen Varianzanalyseverfahren im Chinesischen was die Gestik anbelangt festgestellt werden ($F_{(3,26)}=45,543$; $p<0,001$), ebenso im Deutschen in Bezug auf Gestik ($F_{(3,10)}=12,867$; $p=0,001$), im Spanischen bezüglich Gestik ($F_{(3,9)}=5,814$; $p=0,017$) und Kinesik ($F_{(3,9)}=7,964$; $p=0,007$) und im Portugiesisch ebenfalls im Bereich der Gestik ($F_{(3,9)}=58,088$; $p<0,001$). Mimik und Kinesik scheinen größtenteils nicht vom Grad der Formalität abzuhängen, Grafiken werden daher nicht dargestellt. Für die chinesischen und deutschen Sendungen gilt, dass in allen nicht-informellen Formaten Gestik nahezu ausgeschlossen ist, wohingegen in informellen Sendungen gestische Zeichen „erlaubt“ sind. In chinesischen Nachrichtensendungen fällt auf, dass Gestik v.a. dann eingesetzt wird, wenn zu dialogischen Sprechsituationen übergeleitet wird, also wenn Interaktion mit Berichterstatern vor Ort initiiert wird. Die Verwendung von Gestik ist damit im Chinesischen möglicherweise nicht nur register-abhängig sondern weist auch einen Zusammenhang mit der Anzahl der Gesprächsteilnehmer auf und wird hauptsächlich in monologischen Situationen ausgeschlossen.

Tabelle 8.4: Korrelationen von Gestik, Mimik und Bewegungen mit Registerebene in den einzelnen Sprachen

Sprache		Gesten/min	Gesichtsausdrücke/min	Bewegungen/min
BP	r	0,698*	0,182	-0,424
	p	0,017	0,667	0,295
CN	r	0,742**	-0,214	-0,147
	p	<0,001	0,409	0,574
DE	r	0,766**	0,414	0,492
	p	0,001	0,141	0,074
EP	r	0,127	-0,116	0,045
	p	0,68	0,705	0,884
ES	r	0,732**	0,019	-0,02
	p	0,007	0,957	0,953

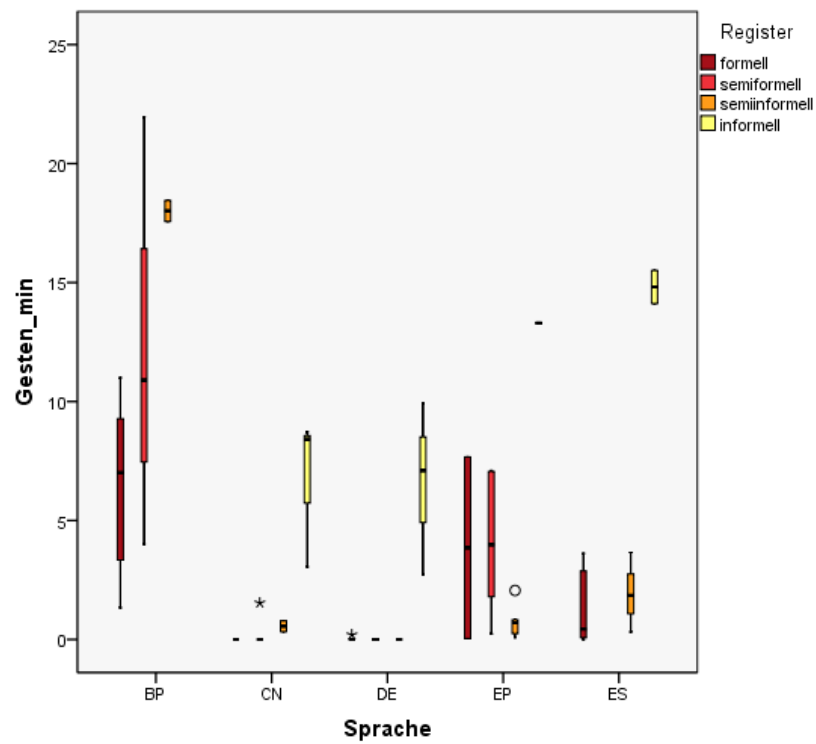


Abbildung 8.4: Verwendung von Gestik in den unterschiedlichen Registerebenen

8.4.2 Qualitative Untersuchung

8.4.2.1 Gestik

8.4.2.1.1 Dauer der Gesten

Nur von brasilianischen Moderatoren werden lange Sequenzen von Gesten produziert, die längere Abschnitte in Äußerungen begleiten. Ein Beispiel hierzu wird in Abbildung 8.5 (A-F) gegeben, in der sechs Bilder aus einer elf Sekunden langen Meldung, die ununterbrochen von Gesten begleitet wird, dargestellt werden. Zwar werden gelegentlich auch in spanischen und deutschen Sendungen mehrere Gesten kurz hintereinander produziert, jedoch lässt sich hier kein fließender Übergang zwischen den einzelnen Bestandteilen ausmachen. Der brasilianische Sprecher (links im Bild) verwendet hierbei beide Arme abwechselnd und beendet die Gestensequenz mit einem Manipulator (Terminologie nach Ekman/Friesen, 1969), indem er eine Textstelle auf dem vor ihm liegenden Blatt mit den Fingern fixiert.



A



B



C



D



E



F

Abbildung 8.5: Gestensequenz eines Nachrichtensprechers (BP)

8.4.2.1.2 Gestenraum

Grundsätzlich ist für die Betrachtung des Gestenraumes relevant, ob mit einer oder mit beiden Händen gestikuliert wird. Nachrichtensprecher aller Nationalitäten verwenden Gesten, die mit beiden Händen oder Armen produziert werden. In brasilianischen und spanischen Nachrichtensendungen können solche Gesten etwas häufiger beobachtet werden als in den anderen.

Ein auffallend großer Gestenraum wird von brasilianischen Nachrichtensprechern verwendet (vgl. Abb. 8.5), in den anderen Sprachen werden hauptsächlich Gesten unmittelbar vor dem Oberkörper oder leicht zur Seite versetzt produziert. Abwechselnde Bewegungen zur linken und rechten Seite werden nur in brasilianischen Sendungen

beobachtet. Besonders in deutschen Sendungen ist zu beobachten, dass bei der Gestik die Hände und Arme kaum vom Oberkörper weg bewegt werden.

Werden von den Moderatoren keine redebegleitenden Bewegungen ausgeführt, fallen ebenfalls Unterschiede auf. Spanische Nachrichtensprecher scheinen in Ruheposition einen größeren Raum zu beanspruchen und strecken oft einen oder beide Arme vom Körper weg und setzen dabei die Hände auf der Schreibtischkante auf, vgl. Abbildungen 8.6 – 8.7.



Abbildung 8.6: Beanspruchter Raum in Ruheposition spanischer Nachrichtensprecherin



Abbildung 8.7: Beanspruchter Raum in Ruheposition deutscher Nachrichtensprecherin

8.4.2.1.3 Motivierte Gesten

Typischerweise werden „hier“, „jetzt“ oder inhaltlich ähnliche Begriffe mit Zeigegesten auf die eigene Position, z.B. mit dem Zeigefinger direkt vor dem Körper auf den Boden/Schreibtisch, untermalt. Gelegentlich werden auch Aufzählungen gestisch mit den Fingern begleitet oder inhaltliche Gegenüberstellungen wie oben beschrieben abgebildet. Der Anteil motivierter Gesten beträgt in den Sprachen folgende Werte:

$$7\% \text{ (BP)} < 12\% \text{ (EP)} = 12\% \text{ (DE)} < 15\% \text{ (ES)} < 16\% \text{ (CN)}$$

Nennenswert ist dabei der relativ geringe Anteil in brasilianischen Sendungen, wo Gesten am häufigsten sind. Im Chinesischen dagegen, das sich durch eine seltene Gestenverwendung auszeichnet, liegt der Anteil motivierter Gesten mehr als doppelt so hoch. Dazu ist anzumerken, dass für diese Erhebung nur eine Stichprobe geringen Umfangs gezogen wurde (mind. sechs Sendungen pro Sprache). Die Abbildungen 8.8 – 8.12 zeigen einige Beispiele hierfür.



Abbildung 8.8: Räumlich-zeitliche motivierte Geste von brasilianischem Nachrichtensprecher: „neste ano“ (de. „in diesem Jahr“)



Abbildung 8.9: Von brasilianischem Nachrichtensprecher verwendetes Emblem für „um pouco“ (de. „ein Wenig“)



Abbildung 8.10: Brasilianischer Nachrichtensprecher stellt den Begriff „solta fumaça“ (de. „stößt Rauch aus“) gestisch dar



Abbildung 8.11: Deutsche Sprecherin untermalt den Ausdruck „verschmolzen“



Abbildung 8.12: Chinesischer Sprecher strukturiert Aufzählung (s. unterer Bildrand)

Alle Nachrichtensprecher machen oft ruckartige Bewegungen mit den Händen und/oder Unterarmen, wobei diese ausschlagende oder bogenförmige Bewegungen beschreiben. Diese nicht motivierten Gesten werden zur Hervorhebung bestimmter Satzteile oder

Wörter verwendet („Taktstockgesten“ nach Ekman/Friesen). In den brasilianischen Nachrichtensendungen fällt dazu eine Geste auf, bei der Daumen und Zeigefinger einen Kreis bilden. Dabei werden die restlichen Finger leicht abgestreckt und die Hand oft auf und ab bewegt. In fast allen Fällen wird dabei eine Zahl betont (Zeitpunkt/Zeitraum, Anzahl, Punktestand uvm.). Hier scheint eine Grammatikalisierung im gestischen Bereich vorzuliegen: Obwohl die Geste des gebildeten Kreises keinen transparenten Bezug zum Inhalt aufweist, also nicht motiviert ist, wird sie auffallend oft in einem bestimmten Kontext verwendet. Es handelt sich um ein konkretes Symbol, das jedoch keine feste eigene Bedeutung aufweist. In Gestenwörterbüchern wird das Symbol meist mit der Bedeutung „in Ordnung“ in Verbindung gebracht.²³³ Die Geste kann auch zeitlich vor dem Zahlwort produziert werden.



Abbildung 8.13: Brasilianische Nachrichtensprecherin betont Zahl in „o filho de *onze* anos“ (de. „der [erst] elfjährige Sohn“)

²³³ Z.B. Meo-Zilio/Mejía (1980: 54), für Spanisch; Harrison (1983: 110) bringt die zu einem Kreis geformten Finger wenn die Handfläche zum Körper zeigt mit einer obszönen Beleidigung in Verbindung. Ostermeier (1986: 85) schreibt zu dieser Geste: „The thumb-and-forefinger-in-a-circle gesture, also referred to as the ring in the United States as the OK sign, [...] conveys a positive meaning indicating everything is fine or perfect. To the Japanese, it signifies money. In France, it may mean zero or worthless while in Tunisia it is a threat. On the other hand, in many parts of Europe, it carries the purely descriptive meaning of naming a day of the week – Thursday.“

8.4.2.2 Persönlicher Bezug

Als „persönlicher Bezug“ wird hier bezeichnet, was dazu führt, dass die monologische Sprechsituation der Nachrichtensprecher durchbrochen und eine direkte Kommunikation mit dem Zuschauer hergestellt wird. Dies kann verbal wie nonverbal umgesetzt werden. Der persönliche Bezug zwischen Moderator und Zuschauer wird in brasilianischen Sendungen aller Stilebenen oft durch nonverbale Mittel wie indexikalische Gesten, aber auch durch die Grußformen verstärkt. In den Sendungen der anderen Sprachen sind Zeichen, die einen direkten Bezug zu den Zuschauern herstellen, sehr selten und auf informelle Formate beschränkt, s. hierzu Abbildung 8.14-8.17. Als verbale Mittel zur Herstellung eines direkten Bezuges zu den Zuschauern fällt in brasilianischen Sendungen die Verabschiedung „Eu espero você [logo mais no Jornal...]“ (de. „ich erwarte dich [später in der Sendung...]“) auch in formelleren Sendungen auf. In spanischen und chinesischen Nachrichtensendungen wird ebenfalls die vertraute Anrede (sp. „tú“, cn. 你, „nǐ“) verwendet, allerdings nur in informellen Sendungen. In spanischen Sendungen ist außerdem die Begrüßung „Hola, ¿qué tal?“ üblich, die wörtlich mit „Hallo, wie geht es?“ übersetzt werden kann – auch wenn der zweite Teile selbstverständlich nicht als tatsächliche Frage zu interpretieren ist. Darüber hinaus begleiten in brasilianischen Sendungen Gesten wie in Abbildung 8.16 auch Inhalte, bei denen die Eigengruppe, also inklusive des Nachrichtensprechers angesprochen ist („incluindo o Brasil“, de. „einschließlich Brasilien“). Einige Meldungen wie z.B. Unwetterwarnungen werden mit persönlichem Bezug auf die Zuschauer ausgesprochen.



Abbildung 8.14: Verabschiedung einer brasilianischen Nachrichtensprecherin



Abbildung 8.15: Verabschiedung einer deutschen Nachrichtensprecherin



Abbildung 8.16: Direkter Bezug eines brasilianischen Nachrichtensprechers zum Zuschauer



Abbildung 8.17: Direkter Bezug einer deutschen Nachrichtensprecherin zum Zuschauer

8.4.2.3 Mimik

Mimik wird in den untersuchten Ländern, wie es für Nachrichtensendungen nicht anders zu erwarten ist, möglichst wertfrei eingesetzt, d.h. es werden keine positiven oder negativen Emotionen ausgedrückt, sondern Teile der Äußerung hervorgehoben. Meist geschieht dies durch das Hochziehen der Augenbrauen, was bei allen Nachrichtensprechern auf ähnliche Weise geschieht. Einzelne Sprecher fallen dabei durch seltenere und kürzere Bewegungen auf, d.h. die Augenbrauen werden nicht so weit hochgezogen und es erscheinen keine auffallenden begleitenden Veränderungen z.B. des Mundes. Wierzbicka (1999: 201 f.) interpretiert diesen mimischen Ausdruck wie folgt: Da das Hochziehen der Augenbrauen einer Bewegung entspricht, die ausgeführt wird, wenn eine Person das Blickfeld zu vergrößern versucht, also ikonisch das Bedürfnis, mehr zu sehen

darstellt, versteht sie darunter die Aussage „ich möchte mehr sehen“. Da sie von der Grundbedeutung „ich denke jetzt“ für eine Bewegung der Augenbrauen ausgeht, sieht sie als zugrunde liegende Nachricht für hochgezogene Augenbrauen die Aussage „ich möchte mehr darüber wissen“. Sie verweist auf zusätzliche spezifischere Interpretationen, die u.a. mit Zweifel, Verwunderung, Interesse oder Überraschung zusammenhängen können. Eine weitere wichtige Bedeutungskomponente ist für Wierzbicka die Veränderung des Wissensstandes, also das Erfahren einer Neuigkeit. Die Gesamtbedeutung hochgezogener Augenbrauen bezeichnet sie folgendermaßen:

(8.1)

I know something now
I want to know more (about this)
I'm thinking now
(nach Wierzbicka 1999: 205)

Für die Nachrichtensprecher selbst sind diese Bedeutungskomponenten nicht sinnvoll. Sie entsprechen jedoch den Gedanken und Reaktionen, die von den Empfängern der Nachricht erwartet werden können. Nachrichtensprecher verwenden diese Gesichtsausdrücke, um auszudrücken: „Es ist bemerkenswert, dass...“, oder, in einfacheren Worten, „Ich weiß, dass du [der Empfänger dieser Nachricht] jetzt etwas weißt, dass du mehr darüber wissen willst, dass du darüber nachdenkst“. Es kann hierbei eine Identifizierung der Nachrichtensprecher mit den Zuschauern postuliert werden, oder zumindest eine kommunikative Annäherung an die Position des Zuschauers. Dieser persönliche Bezug zwischen Sender und Empfänger, der über die Mimik hergestellt wird, ist in allen fünf untersuchten Sprachen vorhanden. Möglicherweise ist eine Funktion davon, den monologischen Charakter der Sprechsituation in Nachrichtensendungen abzuschwächen. Es ist fraglich, ob es überhaupt Nachrichtensendungen gibt, in denen diese Verwendung nonverbaler Mittel *nicht* zu beobachten ist.²³⁴

In allen fünf Ländern können zur Begrüßung und Verabschiedung persönlich erscheinende Gesichtsausdrücke beobachtet werden. Dabei wird dem Zuschauer mit beiden Augen „zugezwinkert“, was in den meisten Fällen von einem Lächeln begleitet wird.

²³⁴ Kendon (1980[1977]: 207) geht davon aus, dass über die artikulatorischen Bewegungen hinaus der Körper beim Sprecher nie vollkommen ruht. Das Hochziehen der Augenbrauen geht auch oft mit auffallenden Bewegungen des Mundes einher.

Die Verwendung von Gesichtsausdrücken nimmt tendenziell mit der gestischer Symbole zu, der Korrelationskoeffizient von 0,287 ist signifikant ($p=0,022$; unter Berücksichtigung des gesamten Materials ohne Trennung nach Sprachen).

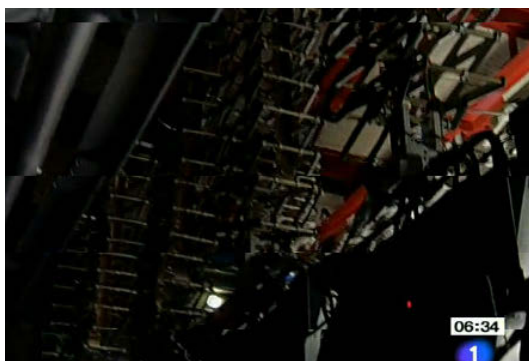
8.4.2.4 Kinesik

Hier ist zunächst anzumerken, dass die Grundvoraussetzungen für den Einsatz von Körperbewegungen als kommunikative Mittel in den Sendungen nicht gleich sind. In deutschen Nachrichtensendungen stehen die Moderatoren für gewöhnlich, wohingegen sie in chinesischen, portugiesischen und spanischen Sendungen sitzen. In brasilianischen Sendungen kommt beides vor. Die Möglichkeiten zur Veränderungen der Körperhaltung sind aufgrund der Beanspruchung unterschiedlicher Muskelgruppen damit nicht vergleichbar. Generell kann festgestellt werden, dass Bewegung in brasilianischen und spanischen Sendungen ein weitaus wichtigeres Konzept darstellt.

Bei allen spanischen und den meisten portugiesischen Nachrichtensprechern, in einem Fall auch bei einem deutschen und in wenigen Fällen bei brasilianischen Nachrichtensprechern, fällt ein Stift auf, der in der überwiegenden Mehrheit der Aufnahmen in der Hand gehalten wird. In vielen Fällen wird dieser Stift (zumindest für deutsche Beobachter) auffallend bewegt. Grundsätzlich kann dies als Symbol für die Professionalität der Moderatoren verstanden werden. Möglicherweise kann das Bewegen des Stiftes auch im Sinne von Dittman (1962) und Ekman/Friesen (1969, 1972) als Ausdruck der allgemeinen Erregung oder Steuern der Aufmerksamkeit interpretiert werden. In deutschen und chinesischen Sendungen scheinen die Blätter in den Händen der Sprecher bzw. in informellen Sendungen Karten mit dem Logo der Sendung als derartige Symbole (für Professionalität) verwendet zu werden. Nur in informellen Sendungen werden diese bewegt und teilweise auch in den Händen zusammengerollt.

Des Weiteren wird das Konzept der Bewegung auch durch Sprecherwechsel dargestellt. Nur eine deutsche Nachrichtensendung („Tagesthemen“), jedoch einige spanische („Telediario matinal“, „Teledario 2“, teilweise „Telediario 1“) sowie die meisten brasilianischen (alle bis auf „Jornal das Dez“ und „Globo Notícia para o G1“) und chinesischen Sendungen (alle bis auf 新闻 1+1 und 东方时空, dōngfāng shíkōng, de. „östlicher Horizont“) werden von zwei Sprechern präsentiert (in den chinesischen Sendungen sind dies in allen Fällen, in brasilianischen in der Regel eine Frau und ein Mann). In spanischen Sendungen werden mindestens die Themenbereiche „Wetter“ und „Sport“ von anderen Nachrichtensprechern präsentiert, wobei hier oft auffällt, was als

nicht traditionelle Rollenverteilung bezeichnet werden kann (Männer präsentieren das Wetter, Frauen die Sportnachrichten, s.u.). Für deutsche Sendungen muss zumindest für alle nicht informellen Nachrichtensendungen als zugrunde liegendes kinesisches Konzept (körperliche) „Unveränderlichkeit“ angenommen werden, da Bewegungen im Vergleich kürzer ausfallen und auch durch die Struktur der Sendung (keine Sprecherwechsel) und den Aufnahmewinkel der Kamera keine Dynamik dargestellt wird. Letzteres trägt in brasilianischen, spanischen und portugiesischen Nachrichtensendungen dazu bei, dass Bewegung dargestellt wird, v.a. in spanischen, aber auch in portugiesischen Nachrichtensendungen werden regelmäßig lange Kamerafahrten durch das Studio gezeigt (vgl. Abb. 8.18 A-F und Abb. 8.19 A-F).



A



B



C



D



E



F

Abbildung 8.18: Kamerafahrt in spanischer Nachrichtensendung



A



B



C



D



E



F

Abbildung 8.19: Kamerafahrt in portugiesischer Nachrichtensendung

Die Häufigkeit der Bewegungen von Kopf/Oberkörper scheint negativ mit der Gestenhäufigkeit ($r=-0,223$) und kaum mit der Häufigkeit mimischer Ausdrücke ($r=0,094$) zu korrelieren und ist in beiden Fällen nicht signifikant (Berücksichtigung des gesamten Materials ohne Trennung nach Sprachen).

8.4.2.5 Weitere Beobachtungen

Mit dem Konzept der Bewegung und damit kinesischen Normen kann auch die Gleichzeitigkeit mehrerer Tonquellen in Verbindung gebracht werden. In spanischen und portugiesischen Nachrichtensendungen wird zumindest während des Nachrichtenüberblicks zu Beginn der Sendung Musik eingespielt, wohingegen in deutschen und chinesischen Sendungen im Normalfall keine Überschneidungen mehrerer Tonquellen vorkommen. In brasilianischen Nachrichtensendungen ist dies selten der Fall, dafür werden häufig Beiträge gesendet, deren Hintergrundgeräusche mindestens den Lautstärkepegel der Stimme des Sprechers erreichen. Zu einem Teil entzieht sich die Regulierung der Tonquellen dabei der technischen Kontrollierbarkeit.

Ein Vergleich der zahlenmäßigen Verhältnisse zwischen den Geschlechtern mit den Werten auf der kulturellen Dimension *Gender Egalitarianism* ergibt deutliche Unterschiede. Am besten ist das Verhältnis in chinesischen Sendungen ausgeglichen, in denen für Gewöhnlich eine Frau und ein Mann abwechselnd die Meldungen lesen, also ein Verhältnis von 1:1 vorliegt. In brasilianischen und deutschen Sendungen ist das Verhältnis nahezu ausgeglichen (Häufigkeit: je 42% bzw. 46% der Nachrichtensprecher sind Frauen in einer Stichprobe von zehn Sendungen pro untersuchter Nachrichtensendung; der Anteil der Nachrichtensprecherinnen die überhaupt in einer Sendung die Nachrichten vortragen liegt jeweils bei 50%). In spanischen Nachrichtensendungen sind Frauen häufiger zu sehen, Anteil und Häufigkeit liegen bei 62%, wobei in weniger formellen Sendungen 80% der Präsentationen der Nachrichten von Frauen durchgeführt werden, die Sportnachrichten werden ebenso in 80% der Fälle von Frauen präsentiert. In portugiesischen Sendungen sind Frauen insgesamt etwas seltener als Männer zu sehen (Häufigkeit: 46%), wobei die formelle Nachrichtensendung nur von Männern präsentiert wird und eine weniger formelle Sendung nur von einer Nachrichtensprecherin; die weiteren Sendungen werden von Frauen und Männern moderiert. In einigen Fällen werden jedoch auf der Internetseite nur die männlichen Nachrichtensprecher vorgestellt und keine Angaben über die Sprecherinnen gemacht, weshalb kein Anteil weiblicher Sprecher angegeben werden kann. Für *Gender Egalitarianism*, definiert als „the degree to which an organization or society minimizes gender role differences while promoting gender equality“ (House et al. 2004: 12),²³⁵

²³⁵ Weiter wird dort ausgeführt, dass in Gesellschaften, die hohe Werte erzielen, Frauen tendenziell mehr Führungspositionen einnehmen, einen gesellschaftlich hohen Status haben, stärker in Entscheidungsprozessen mitwirken, Berufe ausüben, keine rollenspezifischen Berufe ausüben, einen vergleichbaren Bildungsstand haben und eine niedrige Analphabetenquote aufweisen.

werden die Ist-Werte 3,01 (ES) < 3,05 (CN) < 3,1 (DE) < 3,31 (BP) < 3,66 (EP) und die Soll-Werte 3,68 (CN) < 4,82 (ES) < 4,89 (DE) < 4,99 (BP) < 5,13 (EP) angegeben. Portugal geht hieraus als Land hervor, in dem Gleichberechtigung zwischen den Geschlechtern am stärksten angestrebt und am meisten umgesetzt wird, was dem durch die Nachrichtensendungen vermittelten Bild widerspricht. An dieser Stelle soll in Erinnerung gerufen werden, dass im vierten Kapitel geschlechtsspezifische Unterschiede nur für das portugiesische Sprachmaterial festgestellt werden konnten. Im Vergleich fallen also portugiesische Nachrichtensendungen in verschiedenen Aspekten durch Unterschiede zwischen den Geschlechtern auf. Auch für China, das aus der GLOBE-Studie als Gesellschaft mit relativ geringem Interesse an Gleichberechtigung hervorgeht, liefert die Auswertung der Nachrichtensendungen ein anderes Bild.²³⁶ Schließlich ist der Fall Spaniens interessant, das im Vergleich der in der GLOBE-Studie untersuchten Länder wenig Gleichberechtigung umsetzt, aber ein starkes Interesse hieran aufweist. Die stärkere Repräsentation von Frauen in Nachrichtensendungen sowie die themenspezifische Rollenverteilung (s.o.) kann möglicherweise als Bestreben, eine mögliche Benachteiligung auszugleichen, interpretiert werden. Abschließend soll hier festgehalten werden, dass die Untersuchung der Nachrichtensendungen und die GLOBE-Studie aufgrund ihrer nicht ganz vergleichbaren Untersuchungsgegenstände unterschiedliche Ergebnisse liefern.

Kulturelle Unterschiede im proxemischen Verhalten sind auch in den Nachrichtensendungen zu beobachten. In chinesischen Sendungen, in denen eine Frau und ein Mann abwechselnd moderieren, ist in der Regel nur der Sprecher, der gerade vorliest, zu sehen, wohingegen in brasilianischen Sendungen häufig beide im Bild sind. Darüber hinaus kommunizieren Moderatorin und Moderator in brasilianischen und spanischen Sendungen miteinander und können sich in brasilianischen Sendungen berühren, was in chinesischen Sendungen gänzlich ausgeschlossen ist, vgl. Abbildung 8.20.

²³⁶ In der Gesellschaft verbreitete Einstellungen müssen natürlich nicht dem Bild, das ein (staatlicher) Sender vorgibt, entsprechen.



Abbildung 8.20: Proxemisches Verhalten bei Sprecherwechsel in brasilianischer Nachrichtensendung

Abschließend soll hinzugefügt werden, dass Unterschiede im Blickkontakt mit dem Zuschauer auffallen. In chinesischen und deutschen Sendungen wird der Blick gelegentlich von der Kamera abgewandt und auf Unterlagen vor dem Sprecher gerichtet, was in den Sendungen der anderen Länder in der Regel nicht der Fall ist.

8.4.3 Diskussion

Die weitaus deutlicher sichtbare Distanz zwischen Nachrichtensprecher und Zuschauer in chinesischen und deutschen Sendungen ist im verbalen (Anredeformen) wie im nonverbalen Verhalten (hauptsächlich unter Berücksichtigung der Gestik) zu beobachten. Die Ursache liegt möglicherweise darin, dass die (sprachliche) Gestaltung der Nachrichtensendungen mit der Unterscheidung zwischen Fremd- und Eigengruppe zusammenhängt. Für brasilianische Nachrichtensendungen kann angenommen werden, dass sich Nachrichtensprecher und Zuschauer in der gleichen sozialen Gruppe befinden, was bestimmte nähesprachliche Effekte²³⁷ im proxemischen und gestischen Verhalten zur Folge hat. Für chinesische und deutsche Nachrichtensendungen gilt, dass nur in informellen Formaten von Zugehörigkeit der beiden Parteien zur gleichen Gruppe ausgegangen werden kann, was nur in diesen Sendungen zur Verwendung von Gestik und teilweise vertrauter Anrede sowie umgangssprachlichen Formen führt. Eine derartige, aber weniger stark ausgeprägte Differenzierung liegt auch in spanischen Nachrichtensendungen

²³⁷ Der Begriff „Nähesprache“ wird in Anlehnung an Koch/Oesterreicher (1985; 1990) verwendet und beschreibt Kommunikation unter bestimmten Rahmenbedingungen wie Privatheit, Vertrautheit, Emotionalität usw.; vgl. Kap. 9.

vor. Nur die portugiesischen Sendungen lassen diese Interpretation nicht zu; hier kann keine Korrelation von Register und Gestenhäufigkeit beobachtet werden.

Folgende mögliche Interpretationen sollen hier in Erwägung gezogen werden: In manchen Sprachgemeinschaften oder Kulturen, hier im BP, sind nonverbale Kommunikationsmittel stärker in die Verwendung gesprochener Sprache integriert, die Kommunikationskanäle werden ausgeglichener genutzt. Deshalb wird in Nachrichtensendungen aller Register-ebenen mehr gestikuliert als in Sendungen aus anderen Ländern. Da die Untersuchung auf vorbereitetes Sprechen in den Medien eingeschränkt ist, soll eine derartige Pauschalisierung jedoch vermieden werden. Gleichzeitig ist es möglich, dass die Sprech-situation „vorbereitetes Sprechen in den Medien“ Formalität erfordert, die wiederum in den untersuchten Kulturen unterschiedlich ausgeprägt ist. Für die chinesische und die deutsche Kultur mag Wierzbickas kulturelles Skript der persönlichen Autonomie von Bedeutung sein: Formalität erfordert hier Distanz und Reserviertheit zwischen bzw. gegenüber den an der Kommunikationssituation Beteiligten, die sich auch durch einen geringeren Grad an körperlicher Expressivität äußert. Für diese Annahme wird geringere Expressivität mit der Nutzung weniger Kanäle (möglicherweise auch mit einer schwächeren Nutzung aller Kanäle) gleichgesetzt. Eine dritte Erklärung geht davon aus, dass die untersuchte Sprechsituation nicht in allen Kulturen mit Formalität verbunden wird. Dann müssten für chinesische und deutsche Nachrichtensendungen ein formeller Kontext, für brasilianische, portugiesische und spanische Sendungen ein informeller Kontext angenommen werden. Die von Mangold (2003: 859) aufgezählten Merkmale der Nachrichtensprache sind in den deutschen Nachrichtensendungen alle erkennbar. In den Sendungen der romanisch geprägten Kulturen können dagegen Charakteristika festgestellt werden, die Mangold dem *Infotainment*-Stil zuschreibt: Abweichung vom gewöhnlichen Aussagesatz, expressive Wortstellung, Doppelpunkt- bzw. Gedankenstrich-Konstruktionen, Ellipsen und kurzatmiger Satzbau, metaphorische Redensarten, rhetorische Fragen, abgewandelte Buch/Filmtitel.²³⁸ In diesen Kulturen steht damit nicht formelle Vermittlung von Informationen, sondern Unterhaltung im Vordergrund, was die musikalische Untermalung in den Sendungen erklärt. In Anbetracht der Tatsache, dass sich nicht nur chinesische und deutsche Nachrichtensendungen je nach Sendezeit und Inhalt sprachlich unterscheiden (wenn auch deutlicher als die Sendungen der anderen Kulturen), muss davon ausgegangen werden, dass innerhalb der für eine Sprache untersuchten Sendungen ein intendiertes Gefälle bezüglich der Formalität vorliegt. Die überzeugendste Interpretation ist demnach

²³⁸ Manche dieser Merkmale scheinen in jüngerer Zeit auch in der formellen „Tagesschau“ Einzug zu halten.

die zweite, nach der Formalität mit verschiedenen sprachlichen (und gestalterischen) Mitteln umgesetzt wird. Es ist vorstellbar, dass für chinesische und deutsche Nachrichtensprecher Gestik eine emotionale Komponente aufweist und in formellen Sprechsituationen, v.a. in Nachrichtensendungen, ein sachlicher, möglichst neutraler Stil angestrebt wird, mit dem subjektiv erscheinende Mittel wie häufige Gestik als unvereinbar empfunden werden.

Dabei dürfen weitere Unterschiede zwischen den Nachrichtensendungen nicht verschwiegen werden. Hierunter zählen die Hauptsendezeit und Dauer der Sendungen (BP: 20³⁰, 40 Min.; CN: 19⁰⁰, 40 Min.; DE: 20⁰⁰, 15 Min.; ES: 21⁰⁰, 75 Min.; EP: 20⁰⁰, 60 Min.) und die oben erwähnte musikalische Untermalung. Aus der Dauer der Sendungen sowie der musikalischen Untermalung geht hervor, dass nicht alle Sendungen den gleichen Kompromiss aus Information und Unterhaltung anstreben. Daneben scheinen in allen bis auf der deutschen Sendung wirtschaftliche Interessen mittels Werbeunterbrechungen integriert zu werden. Hier soll nicht argumentiert werden, es handele sich um nicht miteinander vergleichbare Situationen, sondern vielmehr, dass eben aufgrund dieser ungleichen Formate kulturelle Unterschiede (z.B. den Tagesablauf betreffende) besonders gut zu beobachten sind. In manchen Gesellschaften geht Formalität in der untersuchten Sprechsituation mit einer Zuordnung der beteiligten Parteien zu unterschiedlichen Gruppen einher (DE, CN), was Distanzsprachmerkmale erfordert, wohingegen in anderen Gesellschaften keine Differenzierung zwischen Eigen- und Fremdgruppe in derselben Sprechsituation unternommen wird (BP, EP, ES) und damit Nähesprachmerkmale zugelassen sind.

Ein Vergleich mit den erhobenen Sprechtempowerten kann aufgrund der geringen Anzahl der untersuchten Sprachen nur unter Vorbehalt angestellt werden. Er zeigt eine Tendenz häufigerer Gestenverwendung in schneller gesprochenen Sprachen: Die Gestenhäufigkeit korreliert positiv mit der Silben- und Wortgeschwindigkeit (SG; s. Tabelle 8.5). Die Korrelationen sind statistisch nicht signifikant. Mit den Individualismuswerten der GLOBE-Studie können keine Zusammenhänge erkannt werden.

Tabelle 8.5: Zusammenhang zwischen Sprechtempo und Häufigkeit nonverbaler Kommunikationsmittel

Tempomaß		Gesten/min	Mimik/min	Bewegungen/min
Laute/s (SG)	r	-0,199	-0,404	0,767
	p	0,748	0,500	0,131
Silben/s (SG)	r	0,701	0,503	-0,240
	p	0,188	0,387	0,697
Morpheme/s	r	-0,001	0,257	0,724
	p	0,999	0,676	0,167
Wörter/s	r	0,654	0,684	-0,442
	p	0,232	0,203	0,456

8.5 Zusammenfassung

Die Untersuchung der Nachrichtensendungen in Hinblick auf die Verwendung nonverbaler Kommunikationsmittel hat ergeben, dass in brasilianischen Sendungen am häufigsten gestikuliert wird, gefolgt von spanischen und portugiesischen Sendungen, und dass in chinesischen und deutschen Sendungen Gestik kaum verwendet wird. Brasilianische Nachrichtensprecher benutzen außerdem einen auffallend großen Gestenraum. In deutschen Sendungen wird auch etwas seltener der Einsatz mimischer Ausdrücke beobachtet, der in brasilianischen und portugiesischen Sendungen am häufigsten stattfindet. Bewegungen des Kopfes und des Oberkörpers sind nur in chinesischen Sendungen häufiger als in den anderen. Die Kameraführung trägt auch zur Umsetzung gewisser kinesischer Konzepte bei. Dabei fällt auf, dass in brasilianischen, besonders aber in spanischen und portugiesischen Nachrichtensendungen Bewegung von Bedeutung ist und in chinesischen und deutschen Sendungen Statik oder Ruhe vermittelt wird. Da nicht davon ausgegangen werden kann, dass die Häufigkeit der Verwendung nonverbaler Kommunikationsmittel vom Sprachsystem (der *langue*) abhängig ist, sondern vielmehr von den arbiträren Konventionen in der Sprachverwendung (der *parole*), kann das Ergebnis als Hinweis darauf interpretiert werden, dass gesellschaftliche Normen diesen Aspekt des Sprechtempos erklären.

Der stärkste Bezug zwischen Nachrichtensprecher und Zuschauer ist in brasilianischen Sendungen zu beobachten. Unter Berücksichtigung auch der Anredeformen wurde in deutschen und chinesischen, aber auch portugiesischen Nachrichtensendungen große Distanz zwischen diesen beiden Diskursteilnehmern festgestellt. Derartige Beobachtungen legen nahe, dass Nachrichtensprecher und Zuschauer in manchen Ländern nicht zur selben Gruppe (der Eigengruppe) zählen und deshalb unterschiedliche Kommunikationsstile die Regel sind.

Anschließend werden alle Ergebnisse in einen Gesamtzusammenhang eingeordnet.

9 Zusammenfassung der Ergebnisse

Im ersten Teil der Arbeit wurde das Sprechtempo unter phonetisch-phonologischen Gesichtspunkten betrachtet. Ziel der Studie war es, die Sprachen BP, Chinesisch, Deutsch, Englisch, EP und Spanisch anhand von normiertem, aber nicht laborspracheähnlichem Sprachmaterial, nämlich Nachrichtensendungen, miteinander zu vergleichen. Dabei wurden verschiedene Parameter des Tempos (Laut-, Silben-, Wort- und Morphemgeschwindigkeit) unter Ein- und Ausschluss der Pausenzeit berücksichtigt. Die Auswertung der erhobenen Daten von je sechs bis acht Sprechern pro Geschlecht und pro Sprache hat ergeben, dass sich die sechs Sprachen signifikant voneinander unterscheiden und keiner der Parameter eine Ausprägung aufweist, die für alle Sprachen als gleich bezeichnet werden kann (im Gegensatz zu Vermutungen in Malécot et al., 1972 u.a.). Gleichzeitig hängt die Einordnung der Sprachen davon ab, welcher Sprechtempoparameter betrachtet wird.

Spanisch kann aufgrund seiner hohen Laut-, Silben- und Wortgeschwindigkeit als schnell bezeichnet werden. Für die Sprachen Deutsch und Englisch ergeben sich eine niedrige Laut-, Silben- und Morphemgeschwindigkeit. BP geht aus diesem Vergleich als mittlere bis schnelle Sprache hervor. Im Chinesischen sind Laut- und Morphemgeschwindigkeit als hoch zu bezeichnen, Silben- und Wortgeschwindigkeit liegen jedoch im mittleren Bereich. Die Einordnung des EP ist deshalb schwierig, weil es die niedrigste Lautgeschwindigkeit, mittlere Silben-, Wort- und Morphemgeschwindigkeit und gleichzeitig hohe Varianzen aufweist. In Bezug auf ihre genetische Verwandtschaft verhalten sich die Sprachen nur was die Morphemgeschwindigkeit anbelangt ähnlich (es sei denn, man zählt das Englische aufgrund des extremen Sprachkontaktes mit dem Französischen zu den romanischen Sprachen – dann ist eine Ähnlichkeit auch bei der Wortgeschwindigkeit zu erkennen). Geschlechtsspezifische Unterschiede im Sprechtempo sind nur im EP statistisch signifikant; hier artikulieren Männer schneller als Frauen.

Zwischen AG und SG bestehen keine nennenswerten Unterschiede, sie korrelieren nahezu perfekt miteinander. Dieser Zusammenhang ist überraschend, da Unterschiede bezüglich der Verwendung von Pausen sehr deutlich ausgeprägt sind, was weiter unten zusammengefasst wird.

Im Sprachvergleich betrachtet konnte als phonetisch-phonologischer Faktor des Sprechtempos die Silbenkomplexität bestätigt werden, die einen starken Einfluss ($r=-0,8$) auf die Silben- und einen schwachen Einfluss auf die Lautgeschwindigkeit ($r=0,4$) ausübt.

Auch die Dauer der artikulatorischen Phrase, definiert als die absolute Zeit zwischen zwei Pausen, hängt negativ mit den Parametern des Tempos zusammen ($r=-0,3$ bis $-0,8$), so dass gilt: Werden mehr Pausen produziert, wird schneller artikuliert. Seltener Pausen, die mit niedrigerem Tempo in Verbindung stehen, weisen auf eine hörerorientierte Strategie in syntaktisch komplexen Sätzen hin.

Die Anzahl der Vokalphoneme korreliert im Sprachvergleich mäßig bis stark negativ mit allen Parametern des Sprechtempos ($r=-0,4$ bis $-0,8$). Sprachen mit einem größeren Vokalsystem neigen zur Nutzung phonologischer Kategorien, die von einem Längenunterschied als sekundärem Merkmal begleitet werden (z.B. gespannte und ungespannte Vokale in den germanischen Sprachen).

Ein höherer Anteil vorne artikulierter Konsonanten hat keine beschleunigende Wirkung auf die Lautgeschwindigkeit ($r=-0,6$). Damit spielt die Agilität der Artikulationsorgane bei der Produktion phonetischer Segmente entgegen der Vermutung keine Rolle für die Schnelligkeit der idealisierten Abfolge der Sprachlaute.

Auch die vermutete das Tempo vermindern Wirkung eines hohen Konsonantenanteils kann nicht verallgemeinert werden: Nur die Silbengeschwindigkeit ($r=-0,6$), nicht jedoch die Lautgeschwindigkeit ($r=0,2$), weist mit dem Konsonantenanteil einen negativen Zusammenhang auf.

Von einer Verallgemeinerung über die untersuchten Sprachen hinaus ist aufgrund des geringen Stichprobenumfangs abzuraten.²³⁹ Lokal gesehen weist die segmentelle Organisation einen Einfluss auf die Lautgeschwindigkeit auf, wie ein Vergleich des Chinesischen mit dem Deutschen zeigt. Da im Chinesischen Konsonantenhäufungen bzw. komplexe Konsonanten, die aus Verschluss, Reibung und Behauchung bestehen, in ihrer Dauer mit einfachen Konsonanten sowohl im Chinesischen als auch im Deutschen vergleichbar sind und entsprechende deutsche Konsonantenhäufungen signifikant länger sind, muss die zeitliche Ausdehnung und Überlappung von Artikulationsbewegungen als Einflussfaktor auf die Lautgeschwindigkeit angenommen werden. Für das Chinesische ist eine Wertung konsonantischer Artikulationsbewegungen in Clustern als komplexe Konsonanten

²³⁹ Die mit den Mittelwerten der Sprachen berechneten Korrelationskoeffizienten sind nicht signifikant; alle Werte für r sind in diesem Kapitel gerundet.

nachvollziehbar. Es weist auch dann noch eine höhere Lautgeschwindigkeit als die anderen Sprachen auf, wenn alle komplexen Konsonanten als ein Segment gewertet werden, so dass die Dauer der Segmente im Chinesischen in jedem Fall geringer ausfällt.

Allgemein zusammengefasst weisen die globalen Sprechtempowerte einen geringen Zusammenhang mit der Beschaffenheit artikulatorischer Elemente auf segmentaler Ebene Segmente auf. Ausschlaggebend ist eher das Zusammenspiel der Segmente und damit auch die phonotaktischen Verhältnisse.

Die Größe sprachlicher Elemente wie das Wort sind für das Sprechtempo ebenfalls von Bedeutung. Darauf deuten die positive Korrelation der Wortlänge mit Laut- und Silbengeschwindigkeit ($r=0,5$ und $0,3$) und die nur schwache negative Korrelation zwischen Wortlänge und Wortgeschwindigkeit ($r=-0,4$) hin, die einen Wortlängeneffekt („Je länger ein Wort, desto schneller wird es produziert“) bestätigt.

Durchschnittswerte sowie Variabilität der Größe (in Lauten gemessen) und Länge (in ms gemessen) von Silben entsprechen nur im ersten Fall den Erwartungen gemäß der Rhythmusphonologie. Die Standardabweichung der Silbendauer fällt im EP und im Englischen am größten aus; EP und Deutsch weisen die höchste Standardabweichung bezüglich der Anzahl von Lauten/Silbe auf.

Deutliche Unterschiede zwischen den Sprachen fallen bei näherer Betrachtung der Pausenstruktur ins Auge. Zwar weisen die Sprachen, vom Spanischen abgesehen, ähnliche Pausenanteile auf (Mittelwert unter Berücksichtigung aller Sprachen: 12,7%). Die geringen Unterschiede von bis zu fünf Prozent sind jedoch statistisch signifikant. Im EP sind hier die Unterschiede zwischen den Geschlechtern statistisch signifikant: In Aufnahmen männlicher Sprecher ist ein höherer Pausenanteil vorhanden.

Ein höherer Pausenanteil geht im Sprachvergleich mit längeren Pausen einher (Anteil und Dauer: $r=0,8$), die aber gleichzeitig seltener sind (Anteil und Abstand zwischen Pausen: $r=-0,2$).

Der Anteil interner Pausen liegt im Chinesischen, im EP und im Spanischen relativ hoch und im Deutschen am niedrigsten. Unterschiede sind statistisch signifikant, wobei sich vier überlappende homogene Untergruppen ergeben.

Gleichzeitig sind Pausendauer und -häufigkeit in den Sprachen sehr ungleich ausgeprägt. Im Deutschen z.B. sind Pausen im Durchschnitt doppelt so lang wie im Chinesischen, wo sie dafür doppelt so häufig sind.

Dehnungen und Häsitationsphänomene werden nur in zwei Sprachen (EP und Spanisch) beobachtet. Bezüglich der Position von Pausen ist zu erwähnen, dass zwischen syntaktisch vollständigen Einheiten hauptsächlich im BP, im Deutschen und im Englischen pausiert wird; innerhalb von syntaktisch vollständigen Einheiten sind in den Sprachen EP, Spanisch und Chinesisch häufig Pausen zu finden. Im Zusammenhang mit dem Pausenverhalten können weiterhin die folgenden Besonderheiten festgestellt werden. Bezüglich des f0-Verlaufs fallen Aneinanderreihungen mit aufzählendem Charakter ohne Deklination auf, die intonatorisch beendeten Einheiten gegenüberstehen. Sie sind ebenfalls im EP und im Spanischen häufig. Was lokale Veränderungen des Tempos betrifft (unter lokal ist hier ein zeitlicher Rahmen von mehreren Wörtern zu verstehen) ist zu erwähnen, dass diese EP und Spanisch als von spontansprachlichen, nächsprachlichen Merkmalen gekennzeichnet erscheinen lassen (s.u.). Das widerlegt teilweise die Überlegungen bezüglich des Sprachmaterials in Hinblick auf die Vergleichbarkeit, und leistet damit einen wichtigen Beitrag zur interkulturellen Forschung.

Für das wahrgenommene Sprechtempo hat sich in einem Wahrnehmungsexperiment mit 75 chinesischen und 76 deutschen Teilnehmern herausgestellt, dass dieses in Abhängigkeit vom phonotaktischen System der Muttersprache der Hörer variiert. Die Silbenkomplexität der Stimuli korreliert mit der mittleren wahrgenommenen Geschwindigkeit deutscher Hörer deutlich stärker ($r=-0,8^{**}$) als mit der chinesischer Hörer ($r=-0,5$), für die wiederum die Lautgeschwindigkeit stärker ($r=0,5$), und v.a. stärker als die Silbengeschwindigkeit ($r=0,4$), von Bedeutung bei der Einschätzung von Stimuli mit unterschiedlichen Silbenkomplexitäten ist. Für deutsche Teilnehmer korreliert die Lautgeschwindigkeit schwächer ($r=0,2$) als die Silbengeschwindigkeit ($r=0,4$) mit der wahrgenommenen Geschwindigkeit. Chinesische Hörer schätzen das Tempo in Abfolgen komplexer Silben höher ein als deutsche, wobei der Einfluss perzipierter Vokalkerne eine Rolle spielt.

Weiterhin konnte in einem informellen Wahrnehmungstest festgestellt werden, dass auch chinesische Versuchsteilnehmer das Tempo romanischer Sprachen höher einschätzen als das germanischer, wobei ein auf Stereotypen beruhender Effekt ausgeschlossen werden kann, da die Versuchsteilnehmer mit keiner Sprache vertraut waren und über diese nicht aufgeklärt wurden. Diese Einschätzung entspricht bisherigen theoretischen Überlegungen und den objektiven Tempowesswerten.

Im zweiten Teil der Arbeit wurden kulturelle Aspekte in den Mittelpunkt der Untersuchungen gestellt. Zunächst konnte in einem Vergleich der erhobenen Sprechtempowerte in den Nachrichtensendungen mit den Werten der GLOBE-Studie

festgestellt werden, dass die langsamer gesprochenen Sprachen mit kulturellen Eigenschaften wie Individualismus, Unsicherheitsvermeidung (*Uncertainty Avoidance*) und langfristig orientiertem Planen (*Future Orientation*) einhergehen. Da dieses Ergebnis teilweise im Widerspruch zu bisherigen Untersuchungen steht, in denen Individualismus allgemein mit höheren Geschwindigkeiten in Verbindung gebracht wird und positive Urteile über Aufnahmen mit hohem Sprechtempo in individualistisch geprägten Kulturen vorliegen, wurde in Aussicht gestellt, dass die Unterscheidung zwischen Eigen- und Fremdgruppe in Interaktion mit dem Grad an Individualismus folgenden Zusammenhang mit dem Sprechtempo aufweist: In individualistisch geprägten Kulturen zählen Nachrichtensprecher und Zuschauer nicht zur selben sozialen Gruppe. Von den Nachrichtensprechern werden daher distanzsprachliche Merkmale verwendet, die möglicherweise ein niedriges Sprechtempo einschließen. In kollektivistisch geprägten Gesellschaften dagegen kann ein höheres Sprechtempo als Begleiterscheinung des Einschlusses des Zuschauers in die Eigengruppe interpretiert werden.

Verbreitete Stereotype zum Sprechtempo romanischer und germanischer Sprachen konnten nicht nur für Sprecher des Deutschen, sondern auch für chinesischsprachige Befragte, die mit keiner der in der Umfrage behandelten Sprachen vertraut sind, bestätigt werden. Die romanischen Sprachen werden ohne Präsentation mit entsprechenden Stimuli von Umfrageteilnehmern als schneller bezeichnet. Die in der phonologisch orientierten Untersuchung gefundenen Ergebnisse entsprechen also auch außerhalb Europas gängigen Stereotypen. Daneben ist in den Fragebögen eine allgemeine Tendenz erkennbar, fremde, nicht gelernte Sprachen als schneller im Vergleich mit der Muttersprache einzuschätzen.

In einem weiteren Wahrnehmungsexperiment beurteilten eine chinesische und eine deutsche Gruppe von Studenten zehn unmanipulierte Ausschnitte aus den untersuchten Nachrichtensendungen bezüglich der Prädikate „vertrauenswürdig“ und „sachlich“. Statistisch signifikante Unterschiede bei der Beurteilung legen nahe, dass für Vertrauenswürdigkeit und Sachlichkeit in den beiden Gruppen unterschiedliche Korrelate vorliegen. Für deutsche Hörer besteht ein negativer Zusammenhang der beiden Prädikate mit Pausenhäufigkeit und Tonhöhenwechsel (r nimmt jeweils einen Wert von ca. $-0,6^{**}$ an). Für chinesische Hörer dagegen ist die Silbengeschwindigkeit von Bedeutung, die ebenfalls stark negativ mit den Beurteilungen korreliert ($r =$ jeweils rund $-0,7^{**}$). Beide Gruppen schätzen Vertrauenswürdigkeit und Sachlichkeit in Aufnahmen der Muttersprache höher ein. Auch in diesem Test geht aus den schriftlichen Kommentaren der Teilnehmer die Bedeutung von Stereotypen für die Beurteilung der Nachrichtensendungen hervor.

Neben der verbalen Sprachproduktion standen nonverbale Kommunikationsmittel im Fokus des zweiten Teils. Es konnte anhand einer weiteren Untersuchung des Materials von insgesamt 81 Nachrichtensprechern gezeigt werden, dass auch hier signifikante Unterschiede zwischen den untersuchten Sprachen bestehen: Von gestischen Zeichen wird im BP deutlich häufiger Gebrauch gemacht als in den anderen Sprachen; im Chinesischen und Deutschen werden Gesten in formellen Nachrichtensendungen gar nicht verwendet. Mimik wird in chinesischen und deutschen Sendungen etwas seltener eingesetzt, in portugiesischen Sendungen dagegen relativ häufig. Chinesische Nachrichtensprecher zeichnen sich durch häufige Bewegungen mit dem Kopf aus. Auch hier erweist sich das Geschlecht der Sprecher nicht als Einflussfaktor.

Dafür kann das Register, das hauptsächlich anhand der Sendezeit in vier Kategorien unterschieden wird, als Faktor der Häufigkeit nonverbaler Kommunikationsmittel bezeichnet werden. Die Korrelation zwischen Register und Gestenhäufigkeit ist in allen Sprachen außer EP (hoch-)signifikant und ist derart ausgeprägt, dass in weniger formellen Sendungen häufiger gestikuliert wird ($r = \text{rund } 0,7^*$ bzw. $0,1$ im EP).

Daneben fallen Unterschiede in der Gestaltung der Nachrichtensendungen auf, v.a. in Bezug auf dargestellte Bewegung (BP, EP, ES) bzw. Statik (CN, DE).

Im Vergleich mit den erhobenen Sprechtempowerten ist eine Tendenz zur häufigeren Verwendung von Gesten in Sprachen mit höherer Silben- und Wortgeschwindigkeit erkennbar ($r=0,7$).

Da die verbalen, v.a. die Anredeformen, sowie die nonverbalen Sprachformen auf niedriger Registerebene, d.h. in informelleren Sendungen, anders ausgeprägt sind, ist davon auszugehen, dass alle formell geprägten Sendungen auch tatsächlich als formell konzipiert sind und nicht bestimmte (z.B. spanische) Nachrichtensendungen informellen Charakter aufweisen, wohingegen sich andere (z.B. deutsche) durch Formalität auszeichnen. Vielmehr muss angenommen werden, dass Formalität in den einzelnen Kulturen unterschiedlich umgesetzt wird und hierdurch die Verwendung verbaler wie nonverbaler Mittel bestimmt wird. Als grundlegender Unterschied zwischen den Nachrichtensendungen der einzelnen Länder wird das Verhältnis zwischen Zuschauer und Nachrichtensprecher angenommen. Sofern sich diese in derselben sozialen Gruppe befinden, werden von den Sprechern nähesprachliche Merkmale wie z.B. häufige Verwendung von Gesten und teilweise weniger formelle Anredeformen verwendet. In den untersuchten Nachrichtensendungen gefundene Merkmale der Nähesprache nach Koch/Oesterreicher (1990) sind z.B. ein niedriger Planungsgrad (vgl. Pausenstruktur und

f0-Verlauf im EP und im Spanischen), die Konjunktion „y“ als Anfangssignal im Spanischen, Dehnungen als Überbrückungsphänomene (frequent im Spanischen) und parataktischer Satzbau (häufig im EP). Für BP können Situationseinbindung, physische Nähe (zwischen den Nachrichtensprechern aber auch als suggerierte Nähe zum Zuschauer) und Präferenz für Gestik als nächsprachliche Merkmale in Nachrichtensendungen erkannt werden. Eine Zuordnung der Zuschauer in die Fremdgruppe scheint in deutschen, chinesischen und womöglich englischen Nachrichtensendungen vorzuliegen und spiegelt sich in Anredeformen und seltener Verwendung nonverbaler Zeichen wider. In chinesischen und deutschen Nachrichtensendungen werden Charakteristika der Sprache der Nähe nur in informellen Sendungen verwendet. Möglicherweise kann auch das Sprechtempo in diese Unterscheidung integriert werden, wobei hier die sprachgemeindeabhängigen Beurteilungen desselben berücksichtigt werden müssen. Hohes Sprechtempo als Zeichen der Inklusion des Hörers in die Wir-Gruppe kann für die brasilianische und spanische Sprachgemeinde angenommen werden. In der deutschen und englischen Sprachgemeinde wird die Zuordnung zur Fremdgruppe durch niedriges Sprechtempo signalisiert. Ob in der chinesischen Sprachgemeinde aufgrund einer andersartigen Beurteilung des Sprechtempos ein hohes Tempo als Signal für die Zuordnung in die Fremdgruppe angenommen werden darf, kann nicht mit Sicherheit gesagt werden. Da sich das EP in Bezug auf die Sprechtempomesswerte nur schwer kategorisieren lässt, ist anhand dieses Merkmals kein Bezug zur Unterscheidung von Eigen- und Fremdgruppe erkennbar.

Die Zugehörigkeit von Nachrichtensprecher und Zuschauer zu unterschiedlichen Gruppen kann auch mit einem intendierten Vorbildcharakter und präskriptiver Funktion der Sprachformen zusammenhängen, wie sie in deutschen und chinesischen Nachrichtensendungen vermutet werden können. Es kann nach wie vor davon ausgegangen werden, dass Hilberts (1986) Vorstellung vom Nachrichtensprecher als kulturelles Vorbild in deutschen Sendungen relevant ist und mehr Formalität fordert, in Nachrichtensendungen anderer Länder dem Nachrichtensprecher jedoch eher die Rolle eines Moderators zukommt, für den in Hilberts Sinne gilt: „Er hat auf lockere Weise seine Hörer auf seiner Welle zu halten und versucht, sprecherisch näher an sie ‘heranzukommen’“ (Hilbert 1986: 163). Deshalb bedienen die Nachrichtensprecher sich (verbaler und nonverbaler) nächsprachlicher Formen.²⁴⁰

²⁴⁰ Von der Zuordnung des Zuschauers zur Eigen- oder Fremdgruppe unabhängig ist die Konzeption der Präsentation der Nachrichtenmeldungen als direkte Kommunikation, die mit einem persönlichen Gespräch zu vergleichen ist, wie Appel (1986: 41) sie fordert, auch mit dem Hinweis, die „Sprache der Mittelschicht so zu

Die Funktion des Fernsehens kann im interkulturellen Vergleich Unterschiede aufweisen, und auch der Umgang mit und das soziale Verhalten während des Fernsehens kann differieren. Es ist vernünftig, davon auszugehen, dass in polychron geprägten Kulturen (s. 7.2.1), in denen Nachrichtensendungen relativ lange dauern, vom Zuschauer nicht wie in monochron geprägten Kulturen, in denen kürzere Sendungen die Regel sind, von Anfang bis Ende verfolgt werden und neben dem informativen auch Unterhaltungscharakter aufweisen.

Wie der Überblick über das Tempo beeinflussende Aspekte im zweiten Kapitel zeigt, sind Pauschalisierungen über standardisierte Sprechsituationen hinaus nicht haltbar. Vor allem individuelle Unterschiede zwischen Nachrichtensprechern dürfen bei diesen Betrachtungen nie außer Acht gelassen werden, selbst wenn die Sprecher nach bestimmten Vorgaben des Senders moderieren. Der Faktor Sprecher sollte daher in zukünftigen Untersuchungen berücksichtigt werden. Daneben sind weitere Untersuchungen anderer standardisierter Sprechsituationen wünschenswert, etwa spontansprachliche formelle Äußerungen (z.B. Interviews) und Gespräche zwischen mehreren Gesprächsteilnehmern. Detailliertere Unterscheidungen von formellen und informellen Kontexten anhand verbaler und nonverbaler Mittel, sowie von einzelsprachlichen Charakteristika, die der Nähe- bzw. Distanzsprache zuzuordnen sind und der situativen Kontexte, in denen diese Varietäten verwendet werden, können für die Fremdsprachendidaktik und interkulturelle Seminare von Vorteil sein.

Literaturverzeichnis

- Abercrombie, David (1967): *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- Almeida, Manuel (1997): Organización temporal del español: el principio de isocronía. *Revista filológica Románica* 14 (1), 29-40.
- Andersen, P. A. / Hecht, M. L. / Hoobler, G. D. / Smallwood, M. (2002): Nonverbal communication across culture. In Gudykunst, W. B. / Mody, B. (Hg.): *Handbook of international and intercultural communication*, 89-106. Thousand Oaks: Sage.
- Antoniadis, Z. / Strube, H. W. (1984): Untersuchungen zur spezifischen Dauer deutscher Vokale. *Phonetica* 41, 72-87.
- Appel, Reinhard (1986): Die Fernsehsprache in den Nachrichten. In: Freyr, Roland Varwig (Hg.): *Sprechkultur im Medienzeitalter*, 33-51. Frankfurt am Main: Scriptor.
- Apple, William / Streeter, Lynn A. / Krauss, Robert M. (1979): Effects of Pitch and Speech Rate on Personal Attributions. *Journal of Personality and Social Psychology* 37 (5), 715-727.
- Arvaniti, Amalia (2009): Rhythm, Timing and the Timing of Rhythm. *Phonetica* 66, 46-63.
- Arvaniti, Amalia / Ross, Tristie (2010): Rhythm classes and speech perception. *Proceedings of Speech Prosody 2010, Chicago, 11-14 May 2010*. Publikation im Internet: <http://speechprosody2010.illinois.edu/papers/100887.pdf> ; Zugriff: 13.09.2011.
- Asu, Eva Liina / Nolan, Francis (2005): Estonian rhythm and the Pairwise Variability Index. *Proceedings of Fonetik 2005, Gothenburg*. Publikation im Internet: <http://www.ling.gu.se/konferenser/fonetik2005/> ; Zugriff: 13.09.2011.
- Atal, B. S. / Chang, J. J. / Mathews, M. V. / Tukey, J. W. (1978): Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *Journal of the Acoustical Society of America* 63 (5), 1535-1555.
- Auer, Peter (1994): Einige Argumente gegen die Silbe als Hauptkategorie. In: Ramers, K. H. / Vater, H. / Woode, H. (Hg.): *Universale Phonologische Strukturen und Prozesse*. Tübingen: Niemeyer.
- Auer, Peter (2001): Silben- und akzentzählende Sprachen. In: Haspelmath, Martin / König, Ekkehard / Oesterreicher, Wulf / Raible, Wolfgang (Hg.): *Language typology and language universals* (2), 1391-1399. Berlin / New York: De Gruyter.
- Auer, Peter / Couper-Kuhlen, Elizabeth (1994): Rhythmus und Tempo konversationeller Alltagssprache. *Zeitschrift für Literaturwissenschaft und Linguistik* 96, 78-106.
- Auer, Peter / Couper-Kuhlen, Elizabeth / Müller, Frank (1999): *Language in Time. The Rhythm and Tempo of Spoken Interaction*. New York / Oxford: Oxford University Press.
- Baayen, R. Harald / Piepenbrock, Richard / Gulikers, L. (1995): *The CELEX lexical Database (CD-ROM)*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania. Zitiert nach Duanmu (2011c: 41).
- Bailey, Charles-James (1980): Evidence for variable syllabic boundaries in English. In: Waugh, Linda R. / van Schooneveld, Cornelis H. (Hg.): *The melody of language*, 25-39. Baltimore: University Park Press.
- Banzina, Elina / Dilley, Laura C. (2010): Context speech rate and duration as cues to native and non-native perception of casually-spoken words in Russian. *Speech Prosody 2010*.

Publikation im Internet: <http://speechprosody2010.illinois.edu/papers/100187.pdf> ; Zugriff: 27.03.2011.

- Ballmer, Thomas T. (1980): The role of pauses and suprasegmentals in a grammar. In: Dechert, Hans W. / Raupach, Manfred (Hg.): *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler*, 211-220. Den Haag (u.a.): Mouton.
- Barbosa, Plínio / Madureira, Sandra (1999): Toward a hierarchical model of rhythm production: evidence from phrase stress domains in Brazilian Portuguese. *14th ICPhS (San Francisco) I*, 297-300.
- Barden, Birgit (1991): Sprechgeschwindigkeit und thematische Struktur. *Arbeitspapier Nr.15 des Projektes Kontextualisierung durch Rhythmus und Intonation*. Universität Konstanz: Fachgruppe Sprachwissenschaft.
- Barik, Henri C. (1977): Cross-linguistic study of temporal characteristics of different types of speech materials. *Language and Speech* 20 (2), 116-126.
- Barnlund, Dean C. (1975): *Public and private self in Japan and the United States*. Ohne Ortsangabe: Intercultural Press.
- Barnlund, Dean C. (1989): *Communicative Styles of Japanese and Americans. Images and Realities*. Belmont, California: Wadsworth Publishing Company.
- Barry, William J. / Andreeva, B. / Russo, M. / Dimitrova, S. / Kostadinova, T. (2003): Do Rhythm Measures tell us anything about Language Type?, *15th ICPhS (Barcelona)*, 2693-2696.
- Barry, William J. / Koreman, Jacques (2009): Do Rhythm Measures Reflect Perceived Rhythm? *Phonetica* 66, 1-17.
- Bassett, Mary R. / O'Connell, Daniel (1978): Pausological Aspects of Guatemalan Children's Narratives. *Bulletin of the Psychonomic Society* 12, 387-389.
- Bassett, Mary R. / O'Connell, Daniel C. / Monahan, William J. (1977): Pausological Aspects of Children's Narratives. *Bulletin of the Psychonomic Society* 9 (3), 166-168.
- Basso, K. H. (1972): 'To give up on words': Silence in Western Apache culture. In: Giglioli, Pier Paolo (Hg.): *Language and Social Context*, 67-86. Harmondsworth: Penguin.
- Bauer, Dominik / Kannampuzha, Jim / Hoole, Phil / Kröger, Bernd J. (2010): Gesture Duration and Articulator Velocity in Plosive-Vowel-Transitions. In: Esposito, A. / Campbell, N. / Vogel, N. / Hussain, A. / Nijholt, A. (Hg.): *Development of Multimodal Interfaces: Active Listening and Synchrony*, LNCS 5967, 346-353. Berlin: Springer.
- Bauer, Laurie (2000): Word. In: Booij, Geert / Lehmann, Christian / Migdan, Joachim (Hg.): *Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung. HSK 17.1*, 247-257. Berlin / New York: Walter de Gruyter.
- Baxter, James C. / Winters, Elaine P. / Hammer, Robert E. (1968): Gestural Behavior During a Brief Interview As a Function of Cognitive Variables. *Journal of Personality and Social Psychology* 8 (3), 303-308.
- Beasley, D. S. / Maki, J. E. (1976): Time- and frequency-altered speech. In: Lass, N. J. (Hg.): *Contemporary issues in experimental phonetics*, 419-458. London: Academic Press.
- Beinhauer, Werner (1930): *Spanische Umgangssprache*. Berlin / Bonn: Dümmler.
- Bell-Berti, Fredericka / Harris, Katherine S. (1981): A Temporal Model of Speech Production. *Phonetica* 38, 9-20.
- Bellugi, Ursula / Fischer, Susan (1972): A comparison of sign language and spoken language. *Cognition* 1, 173-200.

- Benton, Matthew (2010): A Preliminary Analysis of the Relationship of Speech Rate to Speech-Timing Metrics as applied to Large Corpora of Non-Laboratory Speech in English and Chinese Broadcast News. *Speech Prosody 2010*. Publikation im Internet: <http://speechprosody2010.illinois.edu/papers/100423.pdf> ; Zugriff: 15.09.2011.
- Bergmann, Günther (1988): Paralinguale Phänomene. In: Ammon, Ulrich / Dittmar, Norbert / Mattheier, Klaus J. (Hg.): *Sociolinguistics. An International Handbook of the Science of Language and Society. HSK 3.2*, 1216-1227. Berlin / New York: Walter de Gruyter.
- Berschin, Helmut / Fernández-Sevilla, Julio / Felixberger, Josef (2005³): *Die spanische Sprache. Verbreitung, Geschichte, Struktur*. Hildesheim u.a.: Georg Olms Verlag.
- Birdwhistell, Ray L. (1970): *Kinesics and Context. Essays on Body Motion Communication*. Philadelphia: University of Philadelphia Press.
- Bisol, Leda (1999²): *Introdução a estudos de fonologia do português brasileiro*. Porto Alegre: EDIPUCRS.
- Black, J. W. / Tosi, O. / Singh, S. / Takefuta, Y. (1966): A Study of Pauses in Oral Readings of One's Native Language and English. *Language and Speech* 9, 237-241.
- Block, S. / Killen, D. (1996): Speech rates of Australian English-speaking children and adults. *Australian Journal of Human Communication Disorders* 24 (1), 39-44.
- Bloomfield, Leonard (1933): *Language*. New York: Holt.
- Bond, Ronald N. / Feldstein, Stanley (1982): Acoustical Correlates of the Perception of Speech Rate: An Experimental Investigation. *Journal of Psycholinguistic Research* 11 (6), 539-557.
- Boomer, Donald S. (1963): Speech disturbances and bodily movement in interviews. *Journal of Nervous and Mental Disease* 136 (3), 263-266.
- Borzone de Manrique, Ana María / Signorini, Angela (1983): Segmental durations and the rhythm in Spanish. *Journal of Phonetics* 11, 117-128.
- Bose, Ines (1994): *Zur temporalen Struktur frei gesprochenen Texte*. Frankfurt am Main: Theo Hector.
- Boyer, Jacques (2003): Relationship Between Iconic Gestures and Syllabic Boundaries in Communication. In: Rector, Monica / Poggi, Isabella / Trigo, Nadine (Hg.): *Gestures. Meaning and Use*, 183-186. Porto: Edições Universidade Fernando Pessoa.
- Braun, Angelika / Oba, Reiko (2009): Speaking Tempo in Emotional Speech – a Cross-Cultural Study Using Dubbed Speech. Publikation im Internet: <http://www.dfki.de/paraling07/programme/papers/16.pdf> ; Zugriff: 01.09.2011.
- Brubaker, R. S. (1972): Rate and Pause Characteristics of Oral Reading. *Journal of Psycholinguistic Research* 1 (2), 141-147.
- Butcher, Andrew (1981): *Aspects of the speech pause: phonetic correlates and communicative functions*. *Arbeitsberichte Nr. 15*. Kiel: Institut für Phonetik der Universität Kiel.
- Byrd, Dani (1994): Relations of sex and dialect to reduction. *Speech Communication* 15, 39-54.
- Byrd, Dani / Tan, Cheng Cheng (1996): Saying consonant clustery quickly. *Journal of Phonetics* 24, 263-282.
- Calvert, Donald R. (1992²): *Descriptive Phonetics*. New York/Stuttgart: Thieme.
- Canellada, María Josefa / Kuhlman Madsen, John (1987): *Pronunciación del español. Lengua hablada y literaria*. Madrid: Castalia.

- Canepari, Luciano (2005): *A Handbook of Pronunciation. English, Italian, French, German, Spanish, Portuguese, Russian, Arabic, Hindi, Chinese, Japanese, Esperanto*. München: LINCOM.
- Carpenter, Susan / O'Connell, Daniel C. (1988): More than meets the ear: Some variables affecting pause reports. *Language and Communication* 8, 17-27.
- Carrió i Font, Mar / Ríos Mestre, Antonio (1991): Compensatory shortening in Spanish spontaneous speech. *Proceedings of ESCA (Barcelona) 16*: 1-5. PPOSpSt-1991. Publikation im Internet: http://www.isca-speech.org/archive_open/archive_papers/ppospst/pp91_016.pdf ; Zugriff: 13.09.2011.
- Chafcouloff, Michel / Marchal, Alain (1999): Velopharyngeal coarticulation. In: Hardcastle, W. / Hewlett, N. (Hg.): *Coarticulation*, 69-79. Cambridge: Cambridge University Press.
- Chang, Charles B: (2005). The effect of articulatory reduction on intelligibility at fast speech rates: Word recognition in natural fast speech vs. compressed slow speech. University of California, Berkeley. Publikation im Internet: <http://terpconnect.umd.edu/~cbchang/papers/fastspeech.pdf> ; Zugriff: 15.09.2011.
- Chao, Yuen Ren (1968): *A Grammar of Spoken Chinese*. Berkeley/California: University of California Press.
- Cho, Moon-Hwan (2004): Rhythm typology of Korean speech. *Cognitive Processing* 5, 249-253.
- Choi, John-Dongwok / Keating, Pat (1991): Vowel-to-vowel coarticulation in three Slavic languages. *UCLA Working Papers in Phonetics* 78, 78-86.
- Christianfeld, Nicholas / Schachter, Stanley / Bilous, Francis (1991): Filled Pauses and Gestures: It's Not Coincidence. *Journal of Psycholinguistic Research* 20 (1), 1-10.
- Clumeck, H. (1976): Patterns of soft palate movement in six languages. *Journal of Phonetics* 4, 337-351.
- Cohen, R. L. / Borsoi, D. (1996): The role of gestures in description-communication: A cross-sectional study of aging. *Journal of Nonverbal Behavior* 20, 45-63.
- Colina, Sonia (2009): *Spanish Phonology: a syllabic perspective*. Washington: Georgetown University Press.
- Collinder, Björn (1939): *Das Wort als phonetische Einheit*. Uppsala: Almqvist & Wiksell.
- Comrie, Bernard / Corbett, Greville G. (Hg.) (1993): *The Slavonic Languages*. London / New York: Routledge.
- Condon, W. S. (1976): An analysis of behavioral organization. *Sign Language Studies* 13, 285-318.
- Condon, W. S. / Ogston, W. D. (1966): Sound film analysis of normal and pathological behaviour patterns. *Journal of Psychiatric Research* 143, 338-347.
- Cook, Mark / Lallge, Mansur (1967): *The role of vision in interaction*. Unveröffentlichtes Manuskript. University of Oxford.
- Couper-Kuhlen, Elizabeth (1993): *English Speech Rhythm. Form and Function in Everyday Verbal Interaction*. Amsterdam/Philadelphia: John Benjamins.
- Cristófar Silva, Thaís (2007⁹): *Fonética e Fonologia do Português - Roteiro de Estudos e Guia de Exercícios*. São Paulo: Editora Contexto.
- Cruttenden, Alan (2001⁶): *Gimson's Pronunciation of English*. London: Hodder Arnold.
- Crystal, Thomas / House, Arthur (1990): Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America* 88 (1), 101-112.

- Cuenca Villarín, María Heliadora (2000): *Lenguas de compás acentual y lenguas de compás silábico. Revisión teórica e implicaciones pedagógicas. ELIA I*, 41-54.
- Cummins, Fred (2009a): Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*. 37 (1), 16-28.
- Cummins, Fred (2009b): Rhythm as an affordance for the entrainment of movement. *Phonetica*, 66 (1), 15–28.
- Cummins, Fred / Port, Robert (1998): Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26, 145-171.
- Cunha, Celso / Cintra, Luís F. Lindley (2001³): *Nova Gramática do Português Contemporâneo*. Rio de Janeiro: Nova Fronteira.
- Cutugno, Francesco / Savy, Renata (1999): Correlation between segmental reduction and prosodic features in spontaneous speech: the role of tempo. *14th ICPHS (San Fransisco) I*, 471-474.
- Dankovičová, Jana (1999): Articulation rate variation within the intonation phrase in Czech and English. *14th ICPHS 99 San Francisco I*, 269-272.
- Dankovičová, Jana (2001): *The Linguistic Basis of Articulation Rate Variation in Czech. Forum Phonetikum 71*. Frankfurt am Main: Wissenschaftliche Buchhandlung Theo Hector.
- Dankovičová, Jana (2003): Aspects of affective prosody in speech production. *15th ICPHS (Barcelona)*, 2493-2496.
- Dankovičová, Jana / Dellwo, Volker (2007): Czech rhythm and the rhythm class hypothesis. *16th ICPHS (Saarbrücken)*, 1241-1244.
- Darwin, Charles (1872): *The Expression of the Emotions in Man and Animals*. London: Murray.
- Dauer, Rebecca (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, 51-62.
- Dauer, Rebecca (1987): Phonetic and Phonological Components of Language Rhythm. *11th ICPHS (Tallinn) (5)*, 447-450.
- Davitz, J. R. (1964): *Communications of Emotional Meaning*. New York: McGraw-Hill.
- De Haan, Henry (1977): A speech-rate intelligibility threshold for speeded and time-compressed connected speech. *Attention, Perception & Psychophysics* 22 (4), 366-372.
- De Johnson, Teresa / O'Connell, Daniel / Sabin, Edward J. (1979): Temporal analysis of English and Spanish narratives. *Bulletin of the Psychonomic Society* 13 (6), 347-350.
- Deese, James (1980): Pauses, Prosody, and the Demands of Production in Language. In: Dechert, H. / Raupach, M. (Hg.): *Temporal Variables in Speech*, 69-84. Den Haag: Mouton.
- Deese, James (1984): *Thought into Speech: The Psychology of a Language*. Englewood Cliffs: Prentice Hall.
- Delattre, Pierre (1965): *Comparing the phonetic features of English, French, German and Spanish*. London: Harrap / Heidelberg: Groos.
- Delattre, Pierre (1966): A comparison of syllable length conditioning among languages. *IRAL (International Review of Applied Linguistics in Language Teaching) IV/3*, 183-198.
- Delforge, Ann Marie (2008a): Unstressed Vowel Reduction in Andean Spanish. In: Colantoni, Laure / Steele, Jeffrey (Hg.): *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, 107-124. Somerville, MA: Cascadilla Proceedings Project.

- Delforge, Ann Marie (2008b): Gestural Alignment Constraints and Unstressed Vowel Devoicing in Andean Spanish. In: Chang, Charles B. / Haynie, Hannah J. (Hg.): *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 147-155. Somerville, MA: Cascadia Proceedings Project.
- Dellwo, Volker (2006): Rhythm and Speech Rate: A Variation Coefficient for deltaC. In Karnowski, P. / Szigeti, I. (Hg.): *Language and language-processing*, 231-241. Frankfurt am Main: Peter Lang.
- Dellwo, Volker (2009): Choosing the right rate normalization methods for measurements of speech rhythm. Publikation im Internet: http://www.zora.uzh.ch/45236/1/Dellwo_AISV2010.pdf ; Zugriff: 13.03.2012.
- Dellwo, Volker (2010): *Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence*. Dissertation, Universität Bonn. Publikation im Internet: <http://hss.ulb.uni-bonn.de:90/2010/2003/2003.htm> ; Zugriff: 14.09.2011.
- Dellwo, Volker / Fourcin, Adrian / Abberton, Evelyn (2007): Rhythmical classification of languages based on voice parameters. *16th ICPHS (Saarbrücken)*, 1129-1132.
- Dellwo, Volker / Wagner, Petra (2003): Relations between language rhythm and speech rate. *15th ICPHS (Barcelona)*, 471-474.
- DeLong, Alton J. (1981): Phenomenological Space-Time: Toward an Experiential Relativity. *Science, New Series*, 213 (4508), 681-683.
- Deterding, David (2001): The measurement of rhythm: a comparison of Singapore and British English. *Journal of Phonetics*, 29, 217-230.
- Deterding, David / Nolan, Francis (2007): Aspiration and Voicing of Chinese and English Plosives. *16th ICPHS (Saarbrücken)*, 385-388.
- Dictionary Department of the Institute of Linguistics of the Chinese Academy of Social Sciences (2009⁵): *Modern Chinese Words Dictionary*. Peking: Commercial Press [中国社会科学院语言研究所辞典编辑室编. (2009⁵). *现代汉语词典(第五版)*. 北京: 商务印书馆.].
- Dittman, A. T. (1962): The relationship between body movements and moods in interviews. *Journal of Consulting-Psychology* 26, 480. Zitiert nach Wallbott (2003a: 276).
- Dobrogaev, S. M. (1931): The study of reflex in problems of linguistics. In: Marr, W. A. (Hg.): *Lazykovedenie i Materializm* (2), 105-173. Moskau: State Social Economic Publishing House. Zitiert nach Wallbott (2003a: 276).
- Dogil, Grzegorz (2003): Understanding Prosody. In: Rickheit, Gert / Herrmann, Theo / Deutsch, Werner: *Psycholinguistik. Ein internationales Handbuch. HSK 24*, 544-565. Berlin / New York: Walter de Gruyter.
- Dowden, Bradley (2001): Time. In: *Internet Encyclopedia of Philosophy*. <http://www.iep.utm.edu/time/#H2> ; Zugriff: 18.03.2011.
- Drake, Carolyn / Riess Jones, Mari / Baruch, Clarisse (2000): The Development of Rhythmic Attending in Auditory Sequences: Attunement, Referent Period, Focal Attending. *Cognition* 77 (3): 251-288.
- Drommel, Raimund H. (1974): *Probleme, Methoden und Ergebnisse der Pausenforschung. Berichte* (2). Köln: Institut für Phonetik der Universität zu Köln.
- Duanmu, San (2007²): *The Phonology of Standard Chinese*. New York: Oxford University Press.

- Duanmu, San (2009): *Syllable Structure. The Limits of Variation*. New York: Oxford University Press.
- Duanmu, San (2011a): Chinese syllable structure. In: van Oostendorp, Marc / Ewen, Colin J. / Hume, Elizabeth / Rice, Keren (Hg.): *The Blackwell Companion to Phonology* (5), 2754-2777. Wiley-Blackwell.
- Duanmu, San (2011b): The CVX theory of syllable structure. In: Cairns, Charles / Raimy, Eric (Hg.): *Handbook of the Syllable*, 99-127. Leiden / Boston: Brill.
- Duanmu, San (2011c): Chinese Syllable Structure. Publikation im Internet: <http://www-personal.umich.edu/~duanmu/10ChineseSyllable.pdf> ; Zugriff: 14.09.2011.
- Dudenredaktion (2000²²): *Duden. Die deutsche Rechtschreibung*. Mannheim u.a.: Dudenverlag.
- Duez, Danielle (1982): Silent and non silent pauses in three speech-styles. *Language and Speech* 14, 21-32.
- Duez, Danielle (1993): Acoustic Correlates of Subjective Pauses. *Journal Psycholinguistic Research* 22 (1), 21-39.
- Dufter, Andreas (2003): *Typen sprachrhythmischer Konturbildung*. Tübingen. Niermeyer.
- Dupoux, E. / Kakehi, K. / Hirose, Y. / Pallier, C. / Mehler, J. (1999): Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human perception and performance* 25 (6), 1568-1578.
- Dupoux, Emmanuel / Parlato, Erika / Frota, Sonia / Hirose, Yuki / Peperkamp, Sharon (2011): Where do illusory vowels come from? *Journal of Memory and Language* 64 (3), 199-210.
- Eckert, Penelope / McConnell-Ginet, Sally (1998): Communities of practice: Where language, gender, and power all live. In: Coates, J. (Hg.): *Language and Gender*, 484-494. Oxford: Blackwell.
- Editora Melhoramentos (Hg.) (2009): *Michaelis Moderno Dicionário da Língua Portuguesa*. Online-Wörterbuch: <http://michaelis.uol.com.br/> ; Zugriff: 13.09.2011.
- Efron, David (1972[1941]): *Gesture, race and culture. A tentative study of some of the spatio-temporal and "linguistic" aspects of the gestural behavior of Eastern Jews and Southern Italians in New York City, living under similar as well as different environmental conditions*. Den Haag (u.a.): Mouton.
- Eibl-Eibesfeldt, Irenäus (1999⁸): *Grundriss der vergleichenden Verhaltensforschung*. München: Piper.
- Ekman, Paul (1973): Cross-cultural studies of facial expression. In: Ekman, Paul (Hg.): *Darwin and facial expression: A century of research in review*, 169-222. New York: Academic Press.
- Ekman, Paul / Friesen, Wallace V. (1969): The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* 1, 49-98. Zitiert nach Wallbott (2003b: 581).
- Ekman, Paul / Friesen, Wallace V. (1972): Hand movements. *Journal of Communication* 22, 353-374. Zitiert nach Wallbott (2003a: 276).
- Ekman, Paul / Friesen, Wallace V. / O'Sullivan, M. / Chan, A. / Diacoyanni-Tarlatzis, I. / Heider, K. / Krause, R. / LeCompte, W. A. / Pitcairn, T. / Ricci-Bitti, P. E., et al. (1987): Universals And Cultural Differences In The Judgment Of Facial Expressions of Emotion. *Journal of Personality and Social Psychology* 53, 712-717.
- Enfield, N. J. (2001): Lip pointing. A discussion of form and function with reference to data from Laos. *Gesture* 1 (2), 185-212.

- Escudero, Paola / Boersma, Paul / Schurt Rauber, Andréia / Bio, Ricardo A. H. (2009): A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of the Acoustical Society of America* 126 (3), 1379-1393.
- Fackrell, Justin / Vereecken, Helwijn / Buhmann, Jeska / Martens, Jean-Pierre, Van Coile, Bert (2000): Prosodic Variation with Text Type. *Sixth International Conference on Spoken Language Processing (ICSLP 2000)* 3, 231-234.
- Fant, Gunnar / Kruckenberg, A. / Nord, Lennart (1991): Some observations on tempo and speaking style in Swedish text reading. *ESCA Workshop on the phonetics and phonology of speaking styles, Barcelona*. Publikation im Internet: http://www.isca-speech.org/archive_open/archive_papers/ppospst/pp91_023.pdf ; Zugriff: 13.09.2011.
- Fant, Gunnar / Kruckenberg, Anita / Nord, Lennart (1992): Prediction of Syllable Duration, Speech Rate and Tempo. *ICSLP '92* 1, 667-670.
- Farnetani, Edda (1990): V-C-V Lingual Coarticulation and Its Spatiotemporal Domain. In: Hardcastle, William J. / Marchal, Alain (Hg.): *Speech Production and Speech Motoring*, 93-130. Dordrecht u.a.: Kluwer Academic Publishers.
- Feldstein, Stanley / Bond, Ronald N. (1981): Perception of speech rate as a function of vocal intensity and frequency. *Language and Speech* 24, 385-392.
- Feldstein, Stanley / Dohm, Faith-Anne / Crown, Cynthia L. (1993): Gender as a mediator in the perception of speech rate. *Bulletin of the Psychonomic Society* 31 (6), 521-524.
- Fenk-Oczlon, Gertraud / Fenk, August (2006): Speech Rhythm and Speech Rate in Crosslinguistic Comparison. In: Sun, R. / Miyake, N. (Hg.): *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 2480.
- Fenk, August / Fenk-Oczlon, Gertraud / Fenk, Lisa (2006): Syllable Complexity as a Funtion of Word Complexity. In: Solovyev, V. / Goldberg, V. / Polyakov, V. (Hg.): *Text Processing and Cognitive Technologies 11 (VIIIth International Conference "Cognitive Modeling in Linguistics 2005")*, 324-333. Moskau: MISA. Publikation im Internet: http://wwwu.uni-klu.ac.at/gfenk/ESYLLABLECOMPLEXITY_1.pdf ; Zugriff: 25.02.2012.
- Fernández-Conde Rodríguez, Manuel (2005): *La enseñanza de la cultura en la clase de español de los negocios*. Madrid: Arcos.
- Ferragne, Emmanuelle / Pellegrino, François (2004): A comparative account of the suprasegmental and rhythmic features of British English dialects. Publikation im Internet: http://elisa.ddl.ish-lyon.cnrs.fr/fulltext/Ferragne/Ferragne_2004_midl.pdf ; Zugriff: 13.09.2011.
- Fischer-Jørgensen (1964): Sound duration and place of articulation. *Zeitschrift für Sprachwissenschaft und Kommunikationsforschung* 17, 175-207.
- Flachskampf, Ludwig [1938]: *Spanische Gebärdensprache*. In: Krüger, Reinhold (Hg.) (2001): *H. Urtel – L. Flachskampf – M. Wandruszka. Drei Studien zur Körpersprache der Romanen*, 59-104. Berlin: Weidler.
- Fletcher, Janet (1987): Some micro-effects of tempo changes on timing in French. *11th ICPHS (Tallinn)* (3), 129-133. Tallinn: Academy of Sciences of the Estonian SSR.
- Floyd, James L. (1985): *Listening. A Practical Approach*. Glenview / London: Scott, Foresman & Company.
- Fon, Janice (1999): Speech rate as a reflection of variance and invariance in conceptual planning in storytelling. *14th ICPHS 99 (San Francisco)* 1, 663-666.

- Fónagy, I. / Magdics, K. (1960): Speed of utterance in phrases of different length. *Language and Speech* 3, 179-182.
- Ford, Cecilia E. / Thompson, Sandra A. (1996): Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In: Ochs, Elinor / Schegloff, Emanuel A. / Thompson, Sandra A. (Hg.): *Interaction and grammar*, 134-184. Cambridge: Cambridge University Press.
- Fox, R. A. / Lehiste, I. (1987): Discrimination of duration ratios by native English and Estonian Listeners. *Journal of Phonetics* 15, 349-363.
- Fraisse, Pierre (1982): Rhythm and tempo. In: Deutsch, Diana (Hg.): *The psychology of music*, 149-180. New York: Academic Press.
- Fraisse, Paul (1985): *Psychologie der Zeit. Konditionierung, Wahrnehmung, Kontrolle, Zeitschätzung, Zeitbegriff*. München: Ernst Reinhardt Verlag.
- Freedman, N. (1977): Hands, words and mind: On the structuralization of body movements during discourse and the capacity for verbal representation. In: Freedman, N. / Grand, S. (Hg.): *Communicative structures and psychic structures*, 109-132. New York: Plenum Press. Zitiert nach Wallbott (2003a: 277).
- Freedman, N. / Blass, T. / Rifkin, A. / Quitkin, F. (1973): Body movements and the verbal encoding of aggressive affect. *Journal of Personality and Social Psychology* 26, 73-83. Zitiert nach Wallbott (2003a: 277).
- Frischeisen-Köhler, Ida (1933): *Das persönliche Tempo. Eine erbbiologische Untersuchung*. Leipzig: Thieme.
- Gafos, Adamantios I. / Hoole, Philip / Roon, Kevin / Zeroual, Chakir (2010): Variation in overlap and phonological grammar in Moroccan Arabic clusters. In: Fougeron, Cécile / Kühnert, Barbara / D'Impero, Mariapaola / Vallée, Nathalie (Hg.): *Laboratory Phonology 10*, 656-698. Berlin / New York: DeGruyter.
- Galves, Jesus / Garcia, Jesus / Duarte, Denise / Galves, Charlotte (2002): Sonority As a Basis for Rhythmic Class Discrimination. *Proceedings of Prosody 2002, Aix-en-Provence*. Publikation im Internet: <http://www.ime.usp.br/~galves/artigos/sonority.pdf> ; Zugriff: 13.09.2011.
- Garn-Nunn, Pamela / Lynn, James N. (2004³): *Calvert's Descriptive Phonetics*. New York / Stuttgart: Thieme.
- Gay, Thomas (1977): Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America* 62, 183-193.
- Gay, Thomas (1978): Articulatory Units: Segments or Syllables? *Haskins Laboratories Status Reports* 54, 53-64.
- Gay, Thomas (1981): Mechanisms of the control of speech rate. *Phonetica* 38, 148-158.
- Gay, Thomas / Ushijima, T. / Hirose, H. / Cooper, F. S. (1974): Effect of speaking rate on labial consonant-vowel articulation. *Journal of Phonetics* 3, 129-140.
- Gelabert, María José / Martinell Gifre, Emma (1990): *Diccionario de gestos con sus giros más usuales*. Madrid: Edelsa.
- Gibbon, Fiona / Hardcastle, William / Nicolaidis, Katerina (1993): Temporal and Spatial Aspects of Lingual Coarticulation in /kl/ Sequences: A Cross-Linguistic Investigation. *Language and Speech* 36 (2-3), 261-277.
- Gil, David (1986): A prosodic typology of language. *Folia Linguistica* 20, 165-231.

- Gimson, Alfred Charles (1962): *An Introduction to the Pronunciation of English*. London: Arnold.
- Glukhov, A. A. (1975): Statistical Analysis of Speech Pauses for Romance and Germanic Languages. *Soviet Physics. Acoustics*. 21, 71-72.
- Glück, Helmut (2005³): *Metzler Lexikon Sprache*. Stuttgart / Weimar: Metzler.
- Goldman-Eisler, Frieda (1968): *Psycholinguistics. Experiments in Spontaneous Speech*. London / New York: Academic Press.
- Gósy, Mária (1992): *Speech Perception*. Frankfurt am Main: Hector.
- Gottfried, Terry L. / Miller, Joanne L. / Payton, Paula E. (1990): Effect of Speaking Rate on the Perception of Vowels. *Phonetica* 47, 155-172.
- Grabe, Esther / Low, Ee Ling (2002): Durational Variability in Speech and the Rhythm Class Hypothesis. In: Gussenhoven, Carlos / Warner, Natasha (Hg.): *Papers in Laboratory Phonology* 7, 515-546. Berlin: DeGruyter.
- Graham, Jean Anne / Argyle, Michael (1975): A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology* 10 (1), 57-67.
- Grammer, Karl (2004): Körpersignale in menschlicher Interaktion. In: Posner, Roland / Robering, Klaus / Seboek, Thomas: *Semiotik. Ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur. HSK 13.4*, 3448-3487. Berlin / New York: Walter de Gruyter.
- Grawunder, Bodo / Winter, Sven (2010): Acoustic correlates of politeness: prosodic and voice quality measures in polite and informal speech of Korean and German speakers. *Speech Prosody 2010 (Chicago)*. Publikation im Internet: <http://aune.lpl.univ-aix.fr/~sprogis/sp2010/papers/100316.pdf> ; Zugriff: 03.09.2011.
- Greenberg, Joseph (1959): A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26: 178-94.
- Grégoire, A. (1899): Variation de durée de la syllabe française. *La Parole*. Zitiert nach Hegedüs (1957: 13).
- Gregová, Renáta (2011): The CVX Theory of Syllable. A Single-Slot Analysis of the Initial Consonant Clusters in English and in Slovak. *Bulletin of the Transilvania University of Braşov, Series IV: Philology and Cultural Studies Vol. 4 (53), Nr. 1*, 205-212.
- Gros, J. / Mihelič, F. / Pavešić, N. (1999): Slovenian speech timing at different speaking rates. *11th ICPHS 99 (San Francisco) I*, 261-264.
- Grosjean, François (1980a): Temporal variables within and between languages. In: Dechert, Hans W. / Raupach, Manfred (Hg.): *Towards a Cross-Linguistic Assessment of Speech Production*. Frankfurt am Main: Peter Lang.
- Grosjean, François (1980b): Comparative studies of temporal variables in spoken and sign languages: A short review. In: Dechert, Hans W. / Raupach, Manfred: *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler*, 307-312. Den Haag u.a.: Mouton.
- Grosjean, François / Deschamps, A. (1973): Analyse des variables temporelles du français spontané II. Comparaison du français oral dans la description avec l'anglais (description) et avec le français (interview radiophonique). *Phonetica*, 28 (3-4), 191-226.
- Grunwald, Thomas (1983): *Reduktion und Kompensation als Funktion des Sprechgeschwindigkeit im Deutschen*. Hamburg: Helmut Buske Verlag.
- Guerra, Rafael (1983): Estudio estadístico de la sílaba en español. In: Esgueva, M. / Cantarero, M. (Hg.) *Estudios de fonética I. Collectanea Phonetica VII*, 9-112. Madrid: C.S.I.C. Zitiert nach Quilis (1999: 502).

- Guiot, G. / Hertzog, E. / Rondot, P. / Molina, P. (1961): Arrest or acceleration of speech evoked by thalamic stimulation in the course of stereotaxic procedures for Parkinsonism. *Brain* 84, 363-379.
- Gumperz, John (1972[1968]): The speech community. In: Giglioli, Pier Paolo (Hg.): *Language and Social Context*, 219-231. Harmondsworth: Penguin.
- Gunzburger, Deborah (1995): Acoustic and Perceptual Implications of the Transsexual Voice. *Archives of Sexual Behavior* 24 (3), 339-348.
- Haggan, M. (1973): *Cross linguistic aspects of pausing*. Dissertation. University of London. Zitiert nach Butcher (1981: 223).
- Hakkarainen, Heikki J. (1995): *Phonetik des Deutschen*. München: Wilhelm Fink.
- Hakokari, Jussi / Saarni, Tuomo / Salakoski, Tapio / Isoaho, Jouni / Aaltonen, Olli (2007): Measuring relative articulation rate in Finnish utterances. *16th ICPhS (Saarbrücken)*, 1105-1108.
- Hall, Edward Twitchell (1966): *The Hidden Dimension*. Garden City: Doubleday.
- Hall, Edward Twitchell (1976): *Beyond Culture*. New York (u.a.): Anchor.
- Hall, Edward Twitchell (1983): *The Dance of Life*. Garden City: Anchor / Doubleday.
- Hall, Edward Twitchell / Hall, Mildred Reed (1990): *Understanding cultural differences*. Yarmouth: Intercultural Press.
- Hall, Tracy Allan (2000): *Phonologie. Eine Einführung*. Berlin: DeGruyter.
- Halliday, M. A. K. (2006[1992]): A systemic Interpretation of Peking Syllables. In: Webster, Jonathan J. (Hg.) (2006): *Studies in Chinese language (Volume 8 in the Collected Works of M.A.K. Halliday)*, 294-322. London / New York: Continuum.
- Hammarström, G. (1957): The Romance Languages. In: Kaiser, L. (Hg.): *Manual of Phonetics*, 279-288. Amsterdam: North-Holland Publishing Company.
- Hänni, Rolf (1973): *Sprechpausen und die Planung des Sprechens*. Dissertation. Bern.
- Hanson, Petra (2003): *Prosodic phrasing in spontaneous Swedish. Travaux de l'institut de linguistique de Lund* 43. Dissertation. Lund: Department of Linguistics and Phonetics.
- Harris, Katherine S. (1977): The study of articulatory organization: Some negative progress. In: Sawashima, Masayuki / Cooper, Franklin S. (Hg.): *Dynamic Aspects of Speech Production. Current Results, Emerging Problems, and New Instrumentation*, 71-82. Tokio: University of Tokyo Press.
- Harrison, Phyllis A. (1983): *Behaving Brazilian. A Comparison of Brazilian and North American Social Behavior*. Rowley, Massachusetts: Newbury House Publishers.
- Haspelmath, Martin (2002): *Understanding Morphology*. London: Arnold.
- Hayes, Francis (1966): Gesture. In: *Encyclopedia Americana* 12, 627a-d.
- Hegedüs, Ludwig (1957): Sprechtempoanalysen im Ungarischen. *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft* 10 (1), 8-20.
- Heilmann, Christa M / Künzel, Hermann J. (2001): Pausenverhalten und Sprechtempo beim Erzählen und Vorlesen. Eine phonetisch-sprechwissenschaftliche Studie. In: Mauelshagen, Claudia / Seifert, Jan (Hg.): *Sprache und Text in Theorie und Empirie. Beiträge zur germanistischen Sprachwissenschaft. Festschrift für Wolfgang Brandt*, 170-183. Stuttgart: Franz Steiner.

- Hewlett, Nigel / Rendall, Monica (1998): Rural versus urban accent as an influence on the rate of speech. *Journal of the International Phonetic Association* 28, 63-71.
- Hieke, Adolf E. / Kowal, Sabine / O'Connell, Daniel (1983): The Trouble with "Articulatory" Pauses. *Language and Speech* 26 (3), 203-214.
- Hilbert, Peter M. (1986): Die Aussprache des Nachrichtensprechers im Rundfunk als kulturelles Vorbild. In: Freyr Roland Varwig (Hg.): *Sprechkultur im Medienzeitalter*, 161-174. Frankfurt am Main: Scriptor.
- Hirata, Y. (2004): Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 32, 565-89.
- Hirata, Y. / Whiton, J. (2005): Effects of speaking rate on the single/geminate stop distinction in Japanese. *Journal of the Acoustical Society of America* 118, 1647-1660.
- Hoequist, Charles (1983a): Durational Correlates of Linguistic Rhythm Categories. *Phonetica* 40, 19-31.
- Hoequist, Charles (1983b): Syllable Duration in Stress-, Syllable- and Mora-Timed Languages. *Phonetica* 40: 203-237.
- Hoequist, Charles E. Jr. / Kohler, Klaus J. (1986a): Summary of Speech Rate Perception Research at Kiel. In: Kohler, Klaus J.: *Arbeitsberichte des Instituts für Phonetik der Universität Kiel* 22, 5-28.
- Hoequist, Charles E. Jr. / Kohler, Klaus J. (1986b): Further Experiments on Speech Rate Perception with Logatomes. In: Kohler, Klaus J.: *Arbeitsberichte des Instituts für Phonetik der Universität Kiel* 22, 29-136.
- Hofstede, Geert (1980): *Culture's Consequences. International differences in work-related values*. Beverly Hills (u.a.): Sage.
- Hofstede, Geert (2001²): *Culture's Consequences. Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Thousand Oaks (u.a.): Sage.
- Horger, A. (1929): *Általános Fonetika*. Budapest. Zitiert nach Hegedüs (1957: 9).
- Horn, Wilhelm / Lehnert, Martin (1954): *Laut und Leben. Englische Lautgeschichte der neueren Zeit (1400 - 1950) (I)*. Berlin: Deutscher Verlag der Wissenschaften.
- House, R. J./ Hanges, P. J. / Ruiz-Quintanilla, S. A. / Dorfman, P. W. / Javidan, M. / Dickson, M. / Gupta, V. (1999): Cultural influences on leadership and organizations: Project GLOBE. In: Mobley, W. H. (Hg.): *Advances in global leadership (I)*, 171-233. Stanford: JAI Press.
- House, Robert J. / Hanges, Paul J. / Javidan, Mansour / Dorfman, Peter W. / Gupta, Vipin (Hg.) (2004): *Culture, Leadership, and Organizations. The GLOBE Study of 62 Societies*. Thousand Oaks / London / New Delhi: Sage.
- Hudgins, C. V. / Stetson, R. H. (1937): *Relative speed of articulatory movements*. Archives Néerlandaises de Phonétique Expérimentale 13, 85-94.
- Hunold, Cordula (2005): *Chinesische Phonetik: Konzepte, Analysen und Übungsvorschläge für den Unterricht Chinesisch als Fremdsprache*. Bochum: Euroäischer Univesitätsverlag.
- Inhoffen, Nicola (1996): El papel de los gestos en la ordenación y estructuración de la lengua hablada española. In: Kotschi, Thomas / Oesterreicher, Wulf / Zimmermann, Klaus (Hg.): *El español hablado y la cultura oral en España e Hispanoamérica*, 45-68. Frankfurt am Main: Vervuert / Madrid: Iberoamericana.

- Iversen, John R. / Patel, Aniruddh D. / Ohgushi, Kengo (2008): Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustic Society of America* 124 (4), 2263-2271.
- Jacewicz, Ewa / Fox, Robert A. / O'Neill, Caitlin (2009): Articulation rate across dialect, age, and gender. *Language Variation and Change*, 21, 233-256.
- Jang, Tae-Yeoub (2009): Rhythm metrics of spoken Korean. *Language and Linguistics* 46, 169-186.
- Janse, Esther (2003a): Word perception in natural-fast and artificially time-compressed speech. *15th ICPHS (Barcelona)*, 3001-3004.
- Janse, Esther (2003b): *Production and Perception of Fast Speech*. Dissertation. Utrecht University. Publikation im Internet: <http://www.let.uu.nl/~Esther.Janse/personal/dissertatieEJanse.pdf> ; Zugriff: 15.09.2011.
- Jespersen, Otto (1912): *Elementarbuch der Phonetik*. Leipzig: Teubner.
- Jessen, Michael (1998): *Phonetics and Phonology of Tense and Lax Obstruents in German*. Amsterdam / Philadelphia: John Benjamins.
- Johnson, Keith / Martin, Jack (2001): Acoustic Vowel Reduction in Creek: Effects of Distinctive Length and Position in the Word. *Phonetica* 58, 81-102.
- Johnson, T. L. / Strange, W. (1982): Perceptual constancy of vowels in rapid speech. *Journal of the Acoustical Society of America* 72, 1761-1770.
- Kaiser, L. (1939): Biological and statistical research concerning the speech of 216 Dutch students. *Archives Néerlandaises de Phonétique Experimentale* 15, 1-76.
- Kalverkämper, Hartwig (2000): Das Vorurteil in interkulturellen Zeichenprozessen. In: Dietrich, Ute / Winkler, Martina (Hg.): *Okzidentbilder: Konstruktionen und Wahrnehmungen*, 63-82. Leipzig: Leipziger Universitätsverlag.
- Kalverkämper, Hartwig (2003): Nonverbale Kommunikation. In: Ueding, Gert (Hg.): *Historisches Wörterbuch der Rhetorik. Band 6*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Karlgren, H. (1961): Speech Rate and Information Theory. *4th ICPHS (Helsinki)*, 671-677.
- Kasl, S. V. / Mahl, G. F. (1965): The relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology* 1, 425-433.
- Keller, Erich (1990): Speech Motor Timing. In: Hardcastle, William J. / Marchal, Alain (Hg.): *Speech Production and Speech Motoring*, 343-364. Dordrecht u.a.: Kluwer Academic Publishers.
- Keltner, Dacher / Ekman, Paul (2000): Facial Expressions of Emotion. In: Lewis, W. / Haviland-Jones, J. (Hg.): *Handbook of emotions. Second Edition*, 236-249. New York: Guilford Publications.
- Kendall, Tyler (2009): *Speech Rate, Pause, and Linguistic Variation: An Examination Through the Sociolinguistic Archive and Analysis Project*. Dissertation. Durham, NC: Duke University. Publikation im Internet: <http://ncslaap.lib.ncsu.edu/kendall/kendall-dissertation-final.pdf> ; Zugriff: 15.09.2011.
- Kendon, Adam (1980[1977]): Gesticulation and speech: Two aspects of the process of utterance. In: Key, Mary Ritchie (1980) (Hg.): *The Relationship of Verbal and Nonverbal Communication*, 207-227. Den Haag: Mouton.

- Kendon, Adam (1983): Gesture and Speech. How They Interact. In: Wiemann, John M. / Harrison, Randall P. (Hg.): *Nonverbal Interaction (SAGE Annual Reviews of Communication, Volume 11)*, 13-46. Beverly Hills / London / New Delhi: Sage.
- Kendon, Adam (1992): Abstraction in gesture. *Semiotica* 90 (3/4): 225-250.
- Kent, R. / Moll, K. (1972): Cinefluorographix analyses of selected lingual consonants. *Journal of Speech & Hearing Research* 15, 453-473.
- Kent, R.D. (1983): The Segmental Organization of Speech. In: MacNeilage, Peter F. (Hg.): *The Production of Speech*, 57-89. New York u.a.: Springer.
- Koch, Peter / Oesterreicher, Wulf (1985): Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In: *Romanistisches Jahrbuch* 36, 15-43. Berlin / New York: Walter de Gruyter.
- Koch, Peter / Oesterreicher, Wulf (1990): *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*. Tübingen: Max Niemeyer Verlag.
- Kochetov, Alexei / Poulier, Marianne / Son, Minjung (2007): Cross-language differences in overlap and assimilation patterns in Korean and Russian. *16th ICPHS (Saarbrücken)*, 1361-1364.
- Kohler, Klaus J. (1986a): Parameters of speech rate perception in German words and sentences: duration, f0 movement, and f0 level. *Language and Speech* 29, 115-139.
- Kohler, Klaus J. (1986b): Parameters of Speech Rate Perception in German Words and Sentences: Duration, Pitch Movement and Pitch Level. In: Kohler, Klaus J.: *Arbeitsberichte des Instituts für Phonetik der Universität Kiel* 22, 137-178.
- Kohler, Klaus J. (1990): Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In: Hardcastle, W. J. / Marchal, A. (Hg.): *Speech Production and Speech Modelling*, 69-92. Dordrecht u.a.: Kluwer.
- Kohler, Klaus J. (1995²): *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.
- Kohler, Klaus J. / Schäfer, Kurt / Thon, Werner / Timmermann, Gerd (1981): Sprechgeschwindigkeit in Produktion und Perzeption. *Arbeitsberichte* 16, 137-205. Institut für Phonetik, Universität Kiel.
- Koreman, J. / Morland, A. (2007): Speaker consistency of coarticulatory gestures in clusters of labial and velar plosives. *16th ICPHS (Saarbrücken)*. Publikation im Internet: http://www.hf.ntnu.no/isk/koreman/Publications/2007/Poral_CoartGest_finalPREP.pdf ; Zugriff: 15.09.2011.
- Koreman, Jacques (2003): The Perception of Articulation Rate. *15th ICPHS (Barcelona)*, 1711-1714.
- Koreman, Jacques (2006a): Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *Journal of the Acoustic Society of America* 119 (1), 582-596.
- Koreman, Jacques (2006b): The role of articulation rate in distinguishing fast and slow speakers, *Third International Conference on Speech Prosody*. Publikation im Internet: <http://www.hf.ntnu.no/isk/koreman/Publications/2006/speechProsody2006.pdf> ; Zugriff: 15.09.2011.
- Kowal, Sabine (1991): *Über die zeitliche Organisation des Sprechens in der Öffentlichkeit. Pausen, Sprechtempo und Verzögerungen in Interviews und Reden von Politikern*. Bern/Stuttgart/Toronto: Verlag Hans Huber.
- Kowal Sabine / O'Connell, Daniel (1980): Pausological research at Saint Louis University. In: Dechert, Hans W. / Raupach, Manfred (Hg.): *Temporal Variables in Speech. Studies in Honour of Frieda Goldman-Eisler*. Den Haag u.a.: Mouton.

- Kowal, Sabine / O'Connell, Daniel C. (1995): Notation und Transkription in der Gesprächsforschung. *KODIKAS/CODE. Ars Semeiotica: An International Journal of Semiotics* 18, 113-118.
- Kowal, Sabine / O'Connell, Daniel C. (2003): Die Transkription mündlicher Äußerungen. In: Herrmann, Theo / Grabowski, Joachim (Hg): *Sprachproduktion (Enzyklopädie der Psychologie Themenbereich C Theorie und Forschung Serie III Sprache Band I)*, 101-120. Göttingen u.a.: Hogrefe.
- Kowal, Sabine / Wiese, Richard / O'Connell, Daniel C. (1983): The Use of Time in Storytelling. *Language and Speech* 26 (4), 377-392.
- Kozhevnikov, Valerij A. / Chistovich, L. A. (1966): *Speech: articulation and perception. Revised.* Washington, D.C.: U.S Department of Commerce, Joint Publication Research Service.
- Krause, Jean C. / Braid, Louis D. (2002): Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustic Society of America* 112 (5), 2165-2172.
- Kröger, Bernd J. (1996): Zur phonetischen Realisierung von Sprechtempoänderungen unter Einbeziehung von artikulatorischer Reorganisation: Artikulatorische und perzeptive Untersuchungen. In: Gibbon, Daffyd (Hg.): *Natural Language Processing and Speech Technology*, 171-185. Berlin: Mouton de Gruyter.
- Kröger, Bernd J. (1998): *Ein phonetisches Modell der Sprachproduktion*. Tübingen: Niemeyer.
- Kuehn, D. R. / Moll, K. L. (1976): A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics* 4, 303-320.
- Kühnert, Barbara / Nolan, Francis (1999): The Origin of Coarticulation. In: Hardcastle, William J. / Hewlett, Nigel (Hg.): *Coarticulation: Theory, data and techniques*, 7-30. Cambridge: Cambridge University Press.
- LaBarre, W. (1972[1947]): The Cultural Basis of Emotions and Gestures. In: Laver, John / Hucheson, Sandy (Hg.) (1972): *Communication in face to face interaction. Selected readings*, 207-224. Harmondsworth: Penguin.
- Ladefoged, Peter (1983): Cross-Linguistic Studies of Speech Production. In: MacNeilage, Peter F. (Hg.): *The Production of Speech*, 177-188. New York u.a.: Springer.
- Ladefoged, Peter (2005): *Vowels and Consonants. An Introduction to the sounds of languages*. Malden: Blackwell.
- Ladefoged, Peter / Maddieson, Ian (1996): *The sounds of the world's languages*. Cambridge: Blackwell.
- Lallge, Mansur (1972[1971]): *Disfluencies in normal English speech*. Dissertation. University of Oxford, Faculty of Psychological studies. Gelistet in Grammer (2004: 3486) als „Lalleje, Mansur (1971): Disfluencies in normal English speech. Cambridge University“.
- Lallge, Mansur / Cook, Mark (1973): Uncertainty in First Encounters. *Journal of Personality and Social Psychology* 26, 137-141. Gelistet in Grammer (2004: 3486) als „Lalleje, Mansur...“.
- Läuffer, Christiane (1990[1985]): *Some Language-Specific and Universal Aspects of Syllable Structure and Syllabification: Evidence from French and German*. Dissertation. Ann Arbor: University Microfilms International.
- Läuffer, Christiane (1995): Effects of tempo and stress on German syllable structure. *Journal of Linguistics* 31 (2), 227-266.

- Laver, John (1994): *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Lebra, Takie Sugiyama (1987): The cultural significance of silence in Japanese communication. *Multilingua* 6 (4), 343-357.
- Lee, Hyun / Boster, Franklin (1992): Collectivism-Individualism in Perceptions of Speech Rate: A Cross-Cultural Comparison. *Journal of Cross-Cultural Psychology* 23, 377-388.
- Lee, Hyun-Bok (1982): A phonetic study on Korean rhythm. *Malsori [Speech Sounds]* 4, 31-48.
- Lee, Hyun-Bok / Jin, Narn-Taek / Seong, Cheol-Jae / Jung, Il-Jin / Lee, Seung-Mie (1994): An experimental phonetic study of speech rhythm in standard Korean. *Fourth International Conference on Spoken Language ICSLP 1994*, 1091-1094.
- Lee, Hyun-Bok / Seong, Cheol-Jae (1996) Experimental phonetic study of the syllable duration of Korean with respect to the positional effect. *Fourth International Conference on Spoken Language 1996* 2, 1193-1196.
- Lee, Wai-Sum (1999): An Articulatory and Acoustical Analysis of the Syllable-Initial Sibilants and Approximant in Beijing Mandarin. *14th ICPHS 1999 (San Francisco) (1)*, 413-416.
- Leemann, Adrian / Siebenhaar, Beat (2007): Intonational and Temporal features of Swiss German. *16th ICPHS (Saarbrücken)*, 957-960.
- Lehiste, Ilse (1970): *Suprasegmentals*. Cambridge (Massachusetts) / London: M.I.T.
- Lehiste, Ilse (1972): The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America* 6, 2018-2024.
- Lehiste, Ilse (1997): Search for phonetic correlates in Estonian prosody. In: Lehiste, Ilse / Ross, Jaan (Hg.): *Estonian prosody: papers from a symposium*, 11-35. Tallinn: Institute of Estonian Language.
- Lehtonen, Jaakko (1986): Sprechkultur und Sprechnormen: Konsequenzen für die Sprechkultur. In: Varwig, F. R. (Hg.): *Sprechkultur im Medienzeitalter*. Frankfurt: Scriptor.
- Lehtonen, Jaako / Sajavaara, Kari (1985): The Silent Finn. In: Tannen, Deborah / Saville-Troike, Muriel (Hg.): *Perspectives on Silence*, 193-201. Norwood: Ablex Publishing.
- Lemos Monteiro, José (2002⁴): *Morfologia portuguesa*. Campinas: Pontes.
- Lenneberg, Eric H. (1986²): *Biologische Grundlagen der Sprache*. Frankfurt am Main: Suhrkamp.
- Levine, Robert V. / Norenzayan, Ara (1999): The Pace of Life in 31 countries. *Journal of Cross-Cultural Psychology* 30 (2), 178-205.
- Levine, Robert V. / West, Laurie J. / Reis, Harry T. (1980): Perceptions of Time and Punctuality in the United States and Brazil. *Journal of Personality and Social Psychology* 38 (4), 541-550.
- Liberman, A. M. / Cooper, F. S. / Shankweiler, D. P. / Studdert-Kennedy, M. (1967): Perception of the Speech Code. *Psychological Review* 74 (6), 431-461.
- Lin, Hua (2001): Stress and the distribution of the neutral tone in Mandarin. In Xu, Debao (Hg.): *Chinese Phonology in Generative Grammar*, 139-161. San Diego: Academic Press.
- Lin, Hua / Wang, Qian (2007): Mandarin Rhythm: An Acoustic Study. *Journal of Chinese Language and Computing* 17 (3): 127-140.
- Lindblom, B. / Lubker, J. / Gay, T. / Lyberg, B. / Branderud, P. / Holmgren, K. (1987): The Concept of Target and Speech Timing. In: Channon, Robert / Shockey, Linda: *In Honor of Ilse Lehiste*, 161-181. Foris: Dordrecht.

- Lindblom, Bjorn (1963): Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773-1781.
- Lindblom, Björn (1978): Final Lengthening in Speech and Music. In: Gårding, Eva / Bruce, Gösta / Bannert, Robert: *Nordic Prosody. Papers From a Symposium*, 85-101. Lund University: Department of Linguistics.
- Lindblom, Björn (1983): Economy of Speech Gestures. In: MacNeilage, Peter F. (Hg.): *The Production of Speech*, 217-245. New York u.a.: Springer.
- Lipski, John M. (1990): Aspects of Ecuadorian Vowel Reduction. *Hispanic Linguistics* 4 (1), 1-16.
- Liss, Julie M. / White, Laurence / Marrys, Sven L. / Lansford, Kaitlin / Lotto, Andrew J. / Spitzer, Stephanie M. / Caviness, John N. (2009): Quantifying Speech Rhythm Abnormalities in the Dysarthrias. *Journal of Speech, Language and Hearing Research* 52, 1334-1352.
- Lloyd, P. M. / Schnitzer, R. D. (1967): A statistical study of the structure of the Spanish syllable. *Linguistics* 37, 58-72.
- Löfqvist, Anders (1999): Interarticulator phasing, locus equations, and degree of coarticulation. *Journal of the Acoustical Society of America* 106 (4), 2022-2030.
- London, Justin (2004): *Hearing in Time. Psychological Aspects of Musical Meter*. Oxford: Oxford University Press.
- Lope Blanch, Juan M. (1983): *Estudios sobre el español de México*. Mexiko: Universidad Nacional Autónoma de México.
- Loukina, A. / Kochanski, G. / Rosner, B. / Shih, C. / Keane, E. (2011): Rhythm measures and dimensions of durational variation in speech. *Journal of the Acoustical Society of America* 129 (5), 3258-3270.
- Lövgren, Tobias / van Doorn, Jan (2005): Influence of manipulation of short silent pause duration on speech fluency. *Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop*, 123-126.
- Lubker, James / Gay, Thomas (1982): Anticipatory labial coarticulation: experimental, biological, and linguistic variables. *Journal of the Acoustical Society of America* 71 (2), 437-448.
- Luccio, Riccardo (1997): Body behavior as multichannel semiosis. In: Posner, Roland / Robering, Klaus / Seboek, Thomas (Hg.): *Semiotik. Ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur. HSK 13.1*, 345-356. Berlin / New York: Walter de Gruyter.
- Lyberg, B. (1981): *Temporal Properties of spoken Swedish*. Dissertation. Stockholm: Stockholm University.
- Machelett, Kirsten (1996): *Das Lesen von Sonagrammen V1.0. Begleitendes Hypertext-Dokument zur Vorlesung*. Institut für Phonetik und Sprachliche Kommunikation, Universität München. <http://www.phonetik.uni-muenchen.de/studium/skripten/SGL/SGLHome.html> ; Zugriff: 10.03.2012.
- MacLachlan, James (1982): Listener Perception of Time-Compressed Speech... For Radio Commercials. *Journal of Advertising Research* 22 (2), 47-51.
- Maclay, H. / Osgood, C. E. (1959): Hesitation phenomena in spontaneous English speech. *Word* 75, 19-44.
- Magen, Harriet S. (1984): Vowel-to-vowel coarticulation in English and Japanese. *Journal of the Acoustical Society of America* 75 (1), 41.

- Magen, Harriet S. (1997): The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics* 25, 187-205.
- Mahl, G. F. (1956): Disturbances and silences in the patient's speech in psychotherapy. *Journal of Abnormal and Social Psychology* 53, 1-15.
- Malécot, A. / Johnston, R. / Kizziar, P.-A. (1972): Syllabic rate and utterance length in French. *Phonetica* 26, 235-257.
- Mangold, Roland (2003): Massenmedien: Rundfunk und Fernsehen. In: Rickheit, Gert / Herrmann, Theo / Deutsch, Werner: *Psycholinguistik. Ein internationales Handbuch. HSK 24*, 854-863. Berlin / New York: Walter de Gruyter.
- Manuel, Sharon (1987): *Acoustic and perceptual consequences of Vowel-to-Vowel coarticulation in three Bantu languages*. Dissertation. Yale University.
- Manuel, Sharon (1990): The Role of Contrast in Limiting Vowel-to-vowel Coarticulation in Different Languages. *Haskins Laboratories Status Report on Speech Research 103/104*, 1-20.
- Manuel, Sharon (1999): Cross-language studies: Relating language-particular coarticulation patterns to other language-particular facts. In: Hardcastle, William J. / Hewlett, Nigel (Hg.): *Coarticulation. Theory, Data and Techniques*, 179-198. Cambridge: University Press.
- Manuel, Sharon / Krakow, Rena (1984): Universal and Language Particular Aspects of Vowel-to-Vowel Coarticulation. *Haskins Laboratories Status Report on Speech Research 77/78*, 69-87.
- Manusov, Valerie / Jaworski, Adam (2006): Casting Nonverbal Behavior in the Media. Representations and Responses. In: Manusov, Valerie / Patterson, Miles L. (Hg.): *The SAGE Handbook of Nonverbal Communication*, 237-255. Thousand Oaks u.a.: Sage.
- Marín Gálvez, Rafael (1994/1995): La duración vocálica en español. *ELUA (Estudios de Lingüística)* 10, 213-226.
- Marlo, Michael (2004): CVX Theory in CCCCCVX Languages: Implications for Universal Grammar. *Journal of Universal Language* 5, 75-99.
- Martínez Celdrán, E. / Fernández Planas, A. M. (2007): *Manual de Fonética Española*. Barcelona, Ariel.
- Matarazzo, J. D. (1962): Prescribed behavior therapy: suggestions from clinical research. In: Bachrach, A. J. (Hg.): *Experimental Foundations of Clinical Psychology*, 471-509. New York: Basic Books.
- Mateus, Maria Helena / d'Andrade, Ernesto (2000): *The Phonology of Portuguese*. Oxford / New York: Oxford University Press.
- Matsumoto, David (2006): Culture and Nonverbal Behavior. In: Manusov, Valerie / Patterson, Miles L. (Hg.): *The SAGE Handbook of Nonverbal Communication*. Thousand Oaks: SAGE, 219-235.
- Matsumoto, David / Kudoh, Tsutomu / Takeuchi, Sachiko (1996): Changing Patterns of Individualism and Collectivism in the United States and Japan. *Culture Psychology* 2, 77-107.
- Matsumoto, David / Franklin, Brenda / Choi, Jung-Wook / Rodgers, David / Tatani, Haruyo (2002): Cultural Influences on the Expression and Perception of Emotion. In: Gudykunst, William

- B. / Mody, Belly (Hg.): *Handbook of International and Intercultural Communication. Second Edition*. Thousand Oaks: SAGE.
- Matsumoto, David / Yoo, Seung Hee / Fontaine, Johnny (2008): Mapping Expressive Differences Around the World: The Relationship Between Emotional Display Rules and Individualism Versus Collectivism. *Journal of Cross-Cultural Psychology* 39 (55), 55-74.
- Mattheier, Klaus J. (1980): *Pragmatik und Soziologie der Dialekte. Einführung in die kommunikative Dialektologie des Deutschen*. Heidelberg: Quelle & Meyer.
- Mattoso Camara Jr., J. (1972): *The Portuguese Language*. Chicago / London: The University of Chicago Press.
- McAuley, J. Devin (2010): Tempo and rhythm. In: Jones, M. R. / Fay, R. R. / Popper, A. N. (Hg.): *Music Perception. Springer Handbook of Auditory Research* (36), 165-199.
- McClave, Evelyn (1994): Gestural Beats: The Rhythm Hypothesis. *Journal of Psycholinguistic Research* 23 (1), 45-66.
- McNeill, David (1985): So you think gestures are nonverbal? *Psychological Review* 92 (3), 350-371.
- McNeill, David (1992): *Hand and mind: What gesture reveal about thought*. Chicago: Chicago University Press.
- Mehrabian, Albert (1972): *Nonverbal Communication*. Chicago: Aldine-Atherton.
- Mehrabian, Albert (1981²): *Silent Messages: Implicit Communication of Emotions and Attitudes*. Belmont: Wadsworth.
- Meer, Elke van der (2006): Zeit. In: Funke, Joachim / Frensch, Peter A. (Hg.): *Handbuch der Allgemeinen Psychologie – Kognition*, 75-85. Göttingen u.a.: Hogrefe.
- Meinhold, Gottfried (1967): Quantität und Häufigkeit von Pausen in gelesenen deutschen Texten im Zusammenhang mit dem Sprechtempo. *Wissenschaftliche Zeitschrift der Friedrich-Schiller-Universität Jena, Gesellschafts- und sprachwissenschaftliche Reihe* (1), 107-111.
- Menzerath, Paul (1954): *Die Architektonik des deutschen Wortschatzes*. Bonn: Dümmler.
- Menzerath, Paul / de Oleza, J. M. (1928): *Spanische Lautdauer. Eine experimentelle Untersuchung*. Berlin / Leipzig: De Gruyter.
- Meo-Zilio, Giovanni / Mejía, Silvia (1980): *Diccionario de gestos. España e Hispanoamerica. Tomo I A-H*. Bogotá: Imprenta Patriótica del Instituto Caro y Cuervo.
- Meo-Zilio, Giovanni / Mejía, Silvia (1983): *Diccionario de gestos. España e Hispanoamerica. Tomo II I-Z*. Bogotá: Imprenta Patriótica del Instituto Caro y Cuervo.
- Meyer, E. A. / Gombocz, Z. (1909): *Zur Phonetik der ungarischen Sprache*. Uppsala. Zitiert nach Hegedüs (1957: 13).
- Meyer, Rosalee K. / Palmer, Caroline (2001): *Rate and Tactus Effects in Music Performance*. Manuskript. Zitiert nach London (2004: 184).
- Miller, G. A. (1951): Speech and Language. In: Stevens, S. S.: *Handbook of Experimental Psychology*, 789-810. New York: John Wiley.
- Miller, G. A. / Taylor, W. G. (1948): The perception of repeated bursts of noise. *Journal of the Acoustical Society of America* 29, 171-182.
- Miller, J. L. / Grosjean, F. / Lomanto, C. (1984): Articulation rate and its variability in spontaneous speech: reanalysis and some implications. *Phonetica* 41, 215-225.

- Miller, Norman / Maruyama, Geoffrey / Beaber, Rex Julian / Valone, Keith (1976): Speed of Speech and Persuasion. *Journal of Personality and Social Psychology* 34 (4), 615-624.
- Mishima, J. (1951/1952): Fundamental research on the constancy of "mental tempo". *Japanese Journal of Psychology* 22, 27-28.
- Mok, Peggy (2009): On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics* 2, 148-154.
- Mok, Peggy / Lee, Sang Im (2008): Korean speech rhythm using rhythmic measures. *The 18th International Congress of Linguists (CIL18)*. Seoul, Korea. Publikation im Internet: http://cuhk.edu.hk/lin/people/peggy/Mok&Lee_CIL18.pdf ; Zugriff: 13.09.2011.
- Monroy Casas, Rafael (2008): Fonoestilística de las vocales y de las consonantes del español. Su reflejo en los estilos de habla. In: *Lingüística Española Actual XXX*. Madrid: Arcos Libros.
- Morais Barbosa, Jorge (1994a): *Introdução ao Estudo da Fonologia e Morfologia do Português*. Coimbra, Almedina.
- Morais Barbosa, Jorge (1994b): Português: Fonética e Fonologia. In: Holtus, Günter / Metzeltin, Michael / Schmitt, Christian: *Lexikon der romanistischen Linguistik VI* (2), 130-142.
- Morsbach, Helmut (1988): The Importance of Silence and Stillness in Japanese Nonverbal Communication: A Cross-Cultural Approach. In: Poyatos, Fernando (Hg.): *Cross-Cultural Perspectives in Nonverbal Communication*, 201-216. Toronto u.a.: C. J. Hogrefe.
- Moses, Elbert Raymond (1964): *Phonetics*. Englewood Cliffs: Prentice-Hall.
- Müller, Cornelia (1998): *Redebegleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich*. Berlin: Berlin Verlag Arno Spitz GmbH.
- Munhall, K./ Fowler, C. / Hawkins, S. / Saltzman, E. (1992): "Compensatory shortening" in monosyllables of spoken English. In: *Journal of Phonetics* 20, 225-239.
- Nazzi, T. / Bertoncini, J. / Mehler, J. (1998): Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance* 24, 756-766.
- Nakane, Ikuko (2007): *Silence in intercultural communication. Perceptions and performance*. Amsterdam u.a.: Benjamins.
- Nespor, Marina (1990): On the rhythm parameter in phonology. In: Roca, Iggy M. (Hg.): *Logical Issues in Language Acquisition*, 157-175. Foris: Dordrecht.
- Nespor, Marina / Shukla, Mohinish / Mehler, Jacques (2011): Stress-times vs. Syllable-timed Languages. In: van Oostendorp, Marc / Ewen, Colin J. / Hume, Elizabeth / Rice, Keren (Hg.): *The Blackwell Companion to Phonology* (2), Chapter 48. Publikation im Internet: http://www.companiontophonology.com/public/tocnode?query=stress-timed&widen=0&result_number=1&fields=content&from=search&id=g9781405184236_chunk_g978140518423650&type=std ; Zugriff: 05.09.2011.
- Nolan, Francis / Asu, Eva Liina (2009): The Pairwise Variability Index and coexisting rhythms in language. *Phonetica* 66, 64-77.
- Noll, Volker (2008): *O português brasileiro: formação e contrastes*. São Paulo: Editora Globo S.A.
- Nooteboom, Sieb G. (1985): A Functional View of Prosodic Timing in Speech. In: Michon, John A. / Jackson, Janet L. (Hg.): *Time, Mind, and Behavior*. Berlin u.a.: Springer, 242-252.
- Nooteboom, Sieb G., / Eefting, W. (1994): Evidence for the adaptive nature of speech on the phrase level and below. *Phonetica* 51, 92-98.

- Noreen, Adolf (1923): *Wissenschaftliche Betrachtung der Sprache*. Halle: Niemeyer.
- Nöth, Winfried (2000²): *Handbuch der Semiotik*. Stuttgart / Weimar: J. B. Metzler.
- O'Connell, Daniel C. (1980): Cross-Linguistic Investigations of Some Temporal Dimensions of Speech. In: *Dechert, H. W. / Raupach, M.* (Hg.): *Toward a Cross-Linguistic Assessment of Speech production*, 23-39. Frankfurt am Main: Lang.
- Öhmann, S. E. G. (1966): Coarticulation in VCV Utterances: Spectrographic Measurements. *Journal of the Acoustic Society of America* 39 (1), 151-168.
- Ohno, Sumio / Fukumiya, Masamichi / Fujisako, Hiroya (1996): Quantitative Analysis of the Local Speech Rate and its Application to Speech Synthesis. *Proceedings of ICSLP '96*. Publikation im Internet: <http://www.asel.udel.edu/icslp/cdrom/vol4/411/a411.pdf> ; Zugriff: 15.09.2011.
- Okada, Hideo (1999): Japanese. In: International Phonetic Association (Hg.): *Handbook of the International Phonetic Association* 117-119. Cambridge: University Press.
- Oksaar, Els (1988): Kulturemtheorie. Ein Beitrag zur Sprachverwendungsforschung. *Berichte aus den Sitzungen der Joachim Jungius-Gesellschaft der Wissenschaften e. V.* 6 (3). Göttingen: Vandenhoeck & Rupprecht.
- Oliveira, Miguel Jr. (2002): The Role of Pause Occurrence and Pause Duration in the Signalling of Narrative Structure. In: Ranchhod, Elisabete / Mamede, Nuno João (Hg.): *Advances in Natural Language Processing (Proceedings of the 3rd International Conference PorTAL)*, 43-51.
- Orr, D. B. / Friedman, H. L. / Williams, J. C. C. (1965): Trainability of listening comprehension of speeded discourse. *Journal of Educational Psychology* 3, 84-95.
- Os, Els den (1985): Perception of Speech Rate of Dutch and Italian Utterances. *Phonetica* 42, 124-134.
- Os, Els den (1988): *Rhythm and tempo of Dutch and Italian. A contrastive study*. Utrecht: Rijksuniversiteit.
- Osser, Harry / Peng, Frederick (1964): A cross cultural study of speech rate. *Language and Speech* 7 (2), 120-125.
- Ostermeier, Terry H. (1986): Sprechnormen – Internationale Vergleiche. In: Freyr Roland Varwig (Hg.): *Sprechkultur im Medienzeitalter*, 83-91. Frankfurt am Main: Scriptor.
- Ostry, David J. / Munhall, Kevin G. (1985): Control of rate and duration of speech movements. *Journal of the Acoustic Society of America* 77 (2), 640-648.
- Paivio, A. (1965): Personality and audience influence. In: Maher, B. A. (Hg.): *Progress in Experimental Personality Research* 2, 127-173. New York: Academic.
- Parlato-Oliveira, Erika / Christophe, Amme / Hirose, Yuki / Dupoux, Emmanuel (2010): Plasticity of illusory vowel perception in Brazilian-Japanese bilinguals. *Journal of the Acoustical Society of America* 127 (6), 3738-3748.
- Pearson (Hg.) (ohne Jahr): *Longman Dictionary of Contemporary English*. Online-Wörterbuch: <http://www.pearsonlongman.com/index.html> ; Zugriff: 13.03.2011.
- Peng, Ying / Zebrowitz, Leslie A. / Lee, Hoon Koo (1993): The Impact of Cultural Background and Cross-Cultural Experience on Impressions of American and Korean Male Speakers. *Journal of Cross-Cultural Psychology* 24, 203-220.

- Perales, Josu / Cenoz, Jasone (1996): Silence, communicative competence and communication strategies in second language acquisition. In: Grabher, Gudrun M. (Hg.): *Semantics of Silences in Linguistics and Literature*, 67-87. Heidelberg: Winter.
- Pfützinger, Hartmut. R. (1998): Local Speech Rate as a Combination of Syllable and Phone Rate. *ICSLP 98 Sydney* (3), 1087-1090.
- Pfützinger, Hartmut. R. (1999): Local Speech Rate Perception in German Speech. *14th ICPHS (San Francisco)* (2), 893-896.
- Pfützinger, Hartmut R. (2001): *Phonetische Analyse der Sprechgeschwindigkeit*. Dissertation. *FIPKM* 38, 117-264.
- Pfützinger, Hartmut R. / Tamashima, M. (2006): Comparing Perceptual Local Speech Rate of German and Japanese Speech. *3rd International Conference on Speech Prosody* (1), 105-108.
- Pickett, J. M. (1980): *The Sounds of Speech Communication. A Primer of Acoustic Phonetics and Speech Perception*. Baltimore: University Park Press.
- Pompino-Marschall, Bernd (2009³): *Einführung in die Phonetik*. Berlin: de Gruyter.
- Port, Robert F. (2007): The problem of speech patterns in time. In: Gaskell, Gareth M. (Hg.): *The Oxford Handbook of Psycholinguistics*, 503-514. Oxford: Oxford University Press.
- Porter, Richard E. / Samovar, Larry A. (1998): Cultural Influences on Emotional Expression: Implications for Intercultural Communication. In: Anderson, Peter A. / Guerrero, Laura K. (Hg.): *Handbook of Communication and Emotion. Research, Theory, Application, and Contexts*, 451-472. San Diego: Academic Press.
- Posner, Michael J. / Nissen, Mary J. / Klein, Raymond M. (1976): Visual Dominance: An Information Processing Account of Its Origins and Significance. In: *Psychological Review* 83, 157-171. Zitiert nach Grammer (2004: 3485).
- Povel, Dirk-Jan (1985): Time, Rhythms and Tension: In Search of the Determinants of Rhythmicity. In: Michon, John. A. / Jackson, Janet L. (Hg.): *Time, Mind, and Behavior*, 215-252. Berlin u.a.: Springer.
- Press, Ian (1986): *Aspects of the phonology of the Slavonic languages*. Amsterdam: Rodopi.
- Priberam Informática (Hg.) (2008): *Dicionário Priberam da Língua Portuguesa*. Online-Wörterbuch: <http://www.priberam.pt/dlpo/> ; Zugriff: 13.03.2011.
- Pulgram, Ernst (1970): *Syllable, word, nexus, cursus*. Den Haag: Mouton.
- Pulleyblank, E. G. (1984): Vowelles Chinese? An application of the three tiered theory of syllable structure. In: Chan, M. K. M. (Hg.): *Proceedings of the sixteenth international conference on Sino-Tibetan languages and linguistics* (2), 568-619. Seattle: University of Washington. Zitiert nach Duanmu (2009: 266).
- Quené, Hugo (2007): On the just noticeable difference for tempo in speech. *Journal of Phonetics* 35, 353-362.
- Quené, Hugo (2008): Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *Journal of the Acoustical Society of America* 123 (2), 1104-1113.
- Quilis, Antonio (1993): *Tratado de fonología y fonética españolas*. Madrid: Gredos.
- Quilis, Antonio (1999²): *Tratado de fonología y fonética españolas*. Madrid: Gredos.
- Quilis, Antonio (2005⁶): *Principios de fonética y fonología españolas*. Madrid: Arcos.

- Quilis, Antonio / Fernández, Joseph A. (1992¹⁴): *Curso de fonética y fonología españolas. Para estudiantes angloamericanos*. Madrid: Consejo Superior de Investigaciones Científicas.
- Qiu, Mingren (1984): Strukturformel der Phonotagmen im Deutschen und im Chinesischen. In: Fluck, Hans-Rüdiger (Hg.): *Kontrastive Linguistik Deutsch-Chinesisch*, 76-95. Heidelberg: Groos.
- Raith, Joachim (2004): Sprachgemeinschaft – Kommunikationsgemeinschaft. In: Ammon, Ulrich / Dittmar, Norbert / Mattheier, Klaus J. / Trudgill, Peter: *Sociolinguistics. An International Handbook of the Science of Language and Society. HSK 3.1*, 146-158. Berlin / New York: Walter de Gruyter.
- Ramig, Lorraine (1983): Effects of physiological aging on speaking and reading rates. *Journal of Communication Disorders* 16, 217-226.
- Ramus, Franck / Nespor, Marina / Mehler, Jacques (1999): Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265-292.
- Rao, Rajiv (2010): Final Lengthening and Pause Duration in Three Dialects of Spanish. In: Ortega-Llebaria, Marta (Hg.): *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*, 69-82.
- Rauscher, F. H. / Krauss, R. M. / Chen, Y. (1996): Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science* 7, 226-231. Zitiert nach Wallbott (2003a: 278).
- Ray, George / Zahn, Christopher (1990): Regional speech rates in the United States: A preliminary analysis. *Communication Research Reports* 7, 34-47.
- Read, Charles / Zhang, Yun-Fei / Nie, Hong Yin / Ding, Bao-Qing (1986): The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition* 24 (1-2), 31-44.
- Real Academia Española (Hg.) (2001²²): *Diccionario de la lengua española*. Online-Wörterbuch: <http://www.rae.es/rae.html> ; Zugriff: 13.09.2011.
- Rebollo-Couto, Leticia (1997): Le rythme en espagnol et en portugais: syllabique ou accentuel. *Travaux de l'Institut de Phonétique de Strasbourg* 27, 63-90.
- Rebollo Couto, Leticia (1998): Strategies rythmique et fluidité de la parole. *Travaux de l'Institut de Phonétique de Strasbourg* 28, 71-107.
- Recasens, Daniel (1987): An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics* 15, 299-312.
- Recasens, Daniel (1999): Lingual coarticulation. In: Hardcastle, W. J. / Hewlett, N. (Hg.): *Coarticulation. Theory, Data and Techniques*, 80-104. Cambridge: Cambridge University Press.
- Rehder, Peter (2006⁵): *Einführung in die slavischen Sprachen*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Reich, Uli (2004): Ritmo, saliencia prosódica y cliticación en portugués y español. In: Meisenburg, Trudel / Selig, Maria (Hg.): *Nouveaux départs en phonologie : les conceptions sub- et suprasegmentales*, 125-137. Tübingen: Narr.
- Restle, David / Vennemann, Theo (2001): Silbenstruktur. In: Haspelmath, Martin / König, Ekkehard / Oesterreicher, Wulf / Raible, Wolfgang (Hg.): *Language typology and language universals (Sprachtypologie und sprachliche Universalien)*, HSK 20.2, 1310-1336. Berlin / New York: De Gruyter.

- Rietveld, Antonius / Gussenhoven, Carlos (1987): Perceived speech rate and intonation. *Journal of Phonetics* 15, 273-285.
- Roach, Peter (1998): Some Languages are Spoken More Quickly than Others (Myth 18). In: Bauer, Laurie / Trudgill, Peter (Hg.): *Language Myths*, 150-158. London: Penguin.
- Robb, Michael P. / Gillon, Gail T. (2007): Speech rates of New Zealand English- and American English-speaking children. *Advances in Speech-Language Pathology* 9 (2), 1-8.
- Robb, Michael P. / MacLagan, Margaret A. / Chen, Yang (2004): Speaking rates of American and New Zealand varieties of English. *Clinical Linguistics & Phonetics* 18 (1), 1-15.
- Rochet, A. P. / Rochet, B. L. (1991): The effect of vowel height on patterns of assimilatory nasality in French and English. *12th ICPhS (Aix-en-Provence)* (3), 54-57.
- Rosenfeld, H. M. (1966): Approval-seeking and approval-inducing functions of verbal and nonverbal responses in the dyad. *Journal of Personality and Social Psychology* 4, 597-605.
- Rothlauf, Jürgen (2009³): *Interkulturelles Management. Mit Beispielen aus Vietnam, China, Japan, Russland und den Golfstaaten*. München: Oldenburg.
- Russo, Michaela / Barry, William J. (2008): Isochrony reconsidered. Objectifying relations between rhythm measures and speech tempo. *Proceedings Speech Prosody 2008, 4th Conference on Speech Prosody, Campinas, Brazil*. Publikation im Internet: <http://aune.lpl.univ-aix.fr/~sprogis/sp2008/papers/id052.pdf> ; Zugriff: 13.09.2011.
- Santi, Serge (2003): Speech/Gestures Relationships as a Manifestation of a Dynamic and Integrative Motor-Perceptual Model of Behavior. In: Rector, Monica / Poggi, Isabella / Trigo, Nadine (Hg.): *Gestures. Meaning and Use*, 177-182. Porto: Edições Universidade Fernando Pessoa.
- Schefflen, A. E. (1972): Body language and the social order: communications as behavioral control. Englewood Cliffs: Prentice-Hall. Zitiert nach Graham/Argyle (1975: 67).
- Schegloff, Emanuel A. (1982): Discourse as an interactional achievement: some uses of "uh huh" and other things that come between sentences. In: Tannen, D. (Hg.): *Analyzing Discourse: Text and Talk*, 71-93. Washington, D.C.: Georgetown University Press.
- Schegloff, Emanuel A. (1996): Turn organization: one intersection of grammar and interaction. In: Ochs, Elinor / Schegloff, Emanuel A. / Thompson, Sandra A. (Hg.): *Interaction and grammar*, 52-133. Cambridge: Cambridge University Press.
- Scherer, Klaus R. (1979a): Voice and speech correlates of perceived social influence. In: Giles, H. / St. Clair, R. (Hg.): *The Social Psychology of Language*, 88-120. London: Blackwell. Zitiert nach Traunmüller (2004: 664).
- Scherer, Klaus R. (1979b): Nonlinguistic vocal indicators of emotion and psychopathology. In: Izard, Carroll E. (Hg.): *Emotions in personality and psychopathology*, 493-529. New York: Plenum Press. Zitiert nach Bergmann (1988: 1225).
- Schiller, Niels O. (2002): From Phonetics to Cognitive Psychology: Psycholinguistics Has It All. In: Braun, Angelika / Masthoff, Herbert R. (Hg.): *Phonetics and its Applications. Festschrift for Jens-Peter Köster on the Occasion of his 60th Birthday*. Stuttgart: Franz Steiner Verlag.
- Schiller, Niels O. / Meyer, Antje S. / Baayen, Harald R. / Mevelt, Willem J. M. (1996): A Comparison of Lexeme and Speech Syllables in Dutch. *Journal of Quantitative Linguistics* 3 (1), 8-28.

- Schüppert, Anja / Gooskens, Charlotte / van Heuven, Vincente (2009): Do Danes speak more quickly than Swedes? Vortrag auf der 5th International Conference on Language Variation in Europe (ICLaVE #5), Kopenhagen, 25.-27.Juni 2009. Publikation im Internet: www.let.rug.nl/~gooskens/pdf/publ_LVEP_2009.pdf; Zugriff: 02.08.2011.
- Schwab, Sandra / Grosjean, François (2004): La perception du débit en langue seconde. *Phonetica* 61, 84-94.
- Schwegler, Armin / Kempff, Jürgen / Ameal-Guerra, Ana (2010⁴): *Fonética y Fonología españolas*. Hoboken: Wiley.
- Scollon, Ron (1998): *Mediated discourse as social interaction*. London: Longman.
- Scollon Ron / Wang Scollon, Suzanne (2002²): *Intercultural Communication. A Discourse Approach*. Malden, MA, u.a.: Blackwell Publishing.
- Selting, Margret (1995): *Prosodie im Gespräch : Aspekte einer internationalen Phonologie der Konversation*. Tübingen: Niemeyer.
- Semjen, Andras / Vorberg, Dirk / Schulze, Hans-Henning (1998): Getting Synchronized with the Metronome: Comparisons Between Phase and Period Correction. *Psychological Research* 61: 44-55.
- Sherzer, Joel (1973): Verbal and nonverbal deixis: The pointed lip gesture among the San Blas Cuna. *Language in Society* 2, 117-131.
- Shibamoto, Janet (1987): The womanly woman: Manipulation of stereotypical and nonstereotypical features of Japanese female speech. In: Philips, Susan U. / Steele, Susan / Tanz, Christine (Hg.): *Language, gender and sex in comparative perspective*, 26-49. Cambridge: University Press.
- Shockey, Linda (1987): Rate and Reduction: Some Preliminary Evidence. In: Channon, Robert / Shockey, Linda (Hg.): *In Honor of Ilse Lehiste*, 217-242. Foris, Dordrecht.
- Shriberg, Elizabeth (1999): Phonetic Consequences of Speech Disfluency. *14th ICPhS (San Francisco) I*, 619-622.
- Shuter, R. (1977): A Field Study of Nonverbal Communication in Germany, Italy, and the United States. *Communication Monographs* 44, 298-305. Zitiert nach Ostermeier (1986: 91).
- Sievers, Eduard (1901⁵): *Grundzüge der Phonetik*. Leipzig: Breitkopf & Härtel.
- Sifianou, Maria (1997): Silence and politeness. In: Jaworski, Adam (Hg.): *Silence. Interdisciplinary Perspectives*, 63-84. Berlin / New York: Mouton de Gruyter.
- Sigurd, Bengt (1971): Maximal rate and minimal duration of repeated syllables. *Papers from the Institute of Linguistics of the University of Stockholm* 3. Stockholm: University of Stockholm.
- Slembek, Edith (1993): Vorüberlegungen zu Sprechtempo und Pausierung in verschiedenen Kulturen. In: Bonner, Maria / Braun, Edith / Fix, Hans (Hg.): *Nachbarschaften. Festschrift für Max Mangold*. Saarbrücken: SDV.
- Smith, Alexandra (1976): The timing of French, with reflections on syllable-timing. *Work in Progress* 9, 97-108. Edinburgh: Department of Linguistics, Edinburgh University.
- Smith, Bruce L. / Brown, Bruce L. / Strong, William J. / Rencher, Alvin C. (1975): Effects of Speech Rate on Personality Perception. *Language and Speech* 18: 145-152.
- Smith, Bruce L. / Wasowicz, Jan / Preston, Judy (1987): Temporal characteristics of the speech of normal elderly adults. *Journal of Speech and Hearing Research* 30, 522-529.

- Snell, Ada L. (1918): *Pause. A study of its nature and its rhythmical function in verse, especially blank verse*. Ann Arbor: The Ann Arbor press.
- Solé, Maria-Josep / Ohala, John J. (2010): What is and what is not under the control of the speaker: Intrinsic vowel duration. In: Fougeron, Cécile / Kühnert, Barbara / D'Impero, Mariapaola / Vallée, Nathalie (Hg.): *Laboratory Phonology 10*, 608-655. Berlin / New York: DeGruyter.
- Spencer, Andrew (1996): *Phonology. Theory and description*. Oxford: Blackwell.
- Stanienda, Gesine Dagmar (2002): *Statistische und linguistische Analysen von Schweigen aus fremdsprachendidaktischer Sicht*. Aachen: Shaker Verlag.
- Stern, William J. (1900): *Über Psychologie der individuellen Differenzen*. Leipzig: Johann Ambrosius Barth.
- Stetson, Raymond Herbert (1928): *Motor Phonetics*. Den Haag: Nijhoff.
- Stetson, Raymond Herbert (1951): *Motor phonetics: A study of speech movements in action*. Amsterdam: North-Holland Publications.
- Stölten, K. / Engstrand, O. (2002): Effects of sex and age in the Arjeplog dialect: a listening test and measurements of preaspiration and VOT. *Fonetik 2002, the XVth Swedish Phonetics Conference. Quarterly Progress and Status Report 44*: 29-32. Stockholm: Department of Speech, Music and Hearing and Centre for Speech Technology, KTH.
- Strangert, Eva (1985): *Swedish Speech Rhythm in a Cross-Language Perspective*. Stockholm: Almqvist & Wiksell International.
- Street, Richard L. Jr. / Brady, Robert M. / Putman, William B. (1983): The Influence of Speech Rate Stereotypes and Rate Similarity on Listeners' Evaluations of Speakers. *Journal of Language and Social Psychology* 2, 37-56.
- Sundberg, J. (1979): Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics* 7, 71-79.
- Szczepaniak, Renata (2007): *Der phonologisch-typologische Wandel des Deutschen von einer Silben- zu einer Wortsprache*. Berlin: DeGruyter.
- Szczepaniak, Renata (2009): Wortsprachliches Deutsch und silbensprachliches Spanisch. Ein phonologisch-typologischer Vergleich. *Estudios Filológicos Alemanes* 17, 251-267.
- Tajima, Keiichi / Zawaydeh, Bushra / Kitahara, Mafayu (1999): A comparative study of speech rhythm in Arabic, English, and Japanese. *14th ICPHS (San Francisco)* 1, 285-288.
- Tajima, Keiichi / Zawaydeh, Bushra A. / Kitahara, Mafuyu / Port, Robert F. (2000): Finding acoustic evidence for speech rhythm across languages: A speech cycling approach. *Journal of the Acoustic Society of Japan* 21 (5), 287-289.
- Thomson, William (1923): *The Rhythm of Speech*. Glasgow: MacLehose, Jackson and co.
- Tiede, Mark / Shattuck-Hufnagel, Stefanie / Johnson, Beth / Ghosh, Satrajit / Matthies, Melanie / Zandipour, Majid / Perkell, Joseph (2007): Gestural phasing in /kt/ sequences contrasting within and cross word contexts. *16th ICPHS (Saarbrücken)*, 521-524.
- Tillmann, Hans G. / Mansell, Phil (1980): *Phonetik. Lautsprachliche Zeichen, Sprachsignale und lautsprachlicher Kommunikationsprozeß*. Stuttgart: Klett-Cotta.
- Tillmann, Hans G. / Pfitzinger, Hartmut R. (2003): Local Speech Rate: Relationships between Articulation and Speech Acoustics. *15th ICPHS (Barcelona)*, 3177-3180.

- Tosi, Oscar Israel (1965): *A method for acoustic segmentation of continuous sound into pauses and signals and measurement of segment durations*. Dissertation. Columbus: Ohio State University.
- Traunmüller, Hartmut (2004): Paralinguale Phänomene. In: Ammon, Ulrich / Dittmar, Norbert / Mattheier, Klaus / Trudgill, Peter (Hg.): *Sociolinguistics. An International Handbook of the Science of Language and Society. HSK 3.1. Zweite Auflage*, 653-665. Berlin / New York: Walter de Gruyter.
- Trouvain, Jürgen (2004): *Tempo variation in speech production*. Dissertation. Phonus 8. Saarbrücken: Institut für Phonetik, Universität des Saarlandes.
- Trubetzkoy, Nikolas S. (1962³): *Grundzüge der Phonologie*. Göttingen: Vandenhoeck & Ruprecht.
- Uchida, Yukiko / Townswend, Sarah S. M. / Markus, Hazel Rose / Bergsieker, Hilary B. (2009): Emotions as Within or Between People? Cultural Variation in Lay Theories of Emotion Expression and Inference. *Personality and Social Psychology Bulletin* 35 (11), 1427-1439.
- Uhmann, Susanne (1992): Contextualizing Relevance: On Some Forms and Functions of Speech Rate Changes in Everyday Conversation. In: Auer, Peter / Di Luzio, Aldo (Hg.): *The Contextualization of Language*, 297-336. Amsterdam / Philadelphia: John Benjamins.
- Uhmann, Susanne (1997): *Grammatische Regeln und konversationelle Strategien. Fallstudien aus Syntax und Phonologie*. Tübingen: Niemeyer.
- Ulbrich, Christiane (2005): *Phonetische Untersuchungen zur Prosodie der Standardvarietäten des Deutschen in der Bundesrepublik Deutschland, in der Schweiz und in Österreich*. Frankfurt am Main: Peter Lang.
- Umeda, N. (1975): Vowel duration in American English. *Journal of the Acoustical Society of America* 58, 434-445.
- Urtel, Hermann (1928): Zur Gebärdensprache. In: *Beiträge zur portugiesischen Volkskunde. Abhandlungen aus dem Gebiet der Altertumskunde, Band 27, Reihe B. Völkerkunde, Kulturgeschichte und Sprachen, Band 15*. Hamburg: Kommissionsverlag L. Friedrichsen & Co. Zitiert nach Flachskampf (2001[1938]: 73).
- Verhoeven, Jo / Pauw, Guy de / Kloots, Hanne (2004): Speech Rate in a Pluricentric Language: A Comparison Between Dutch in Belgium and the Netherlands. *Language and Speech* 47 (3), 297-308.
- Von Essen, Otto (1949): Sprechtempo als Ausdruck psychischen Geschehens. *Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft*, 3. Jahrgang, 317-340. Berlin: Akademie.
- Wachtel, Stefan (2002): *Sprechwissenschaftliche Untersuchungen zum Moderieren im Rundfunkjournalismus*. St. Ingbert: Röhrig Universitätsverlag.
- Wagner, Petra / Dellwo, Volker (2004): Introducing YARD (Yet Another Rhythm Determination) And Re-Introducing Isochrony to Rhythm Research. *Speech Prosody* 2004, 227-230.
- Wallbott, Harald (2003a): Gestik und Mimik beim Sprechen. In: Rickheit, Gert / Herrmann, Theo / Deutsch, Werner (Hg.): *Psycholinguistik. Ein internationales Handbuch. HSK 24*, 262-279. Berlin / New York: Walter de Gruyter.
- Wallbott, Harald (2003b): Nonverbale Komponenten der Sprachproduktion. In: Herrmann, Theo / Grabowski, Joachim (Hg.): *Sprachproduktion. Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie III: Sprache, Band 1*, 561-583. Göttingen u.a.: Hogrefe.

- Walker, M. / Nazmi, M. K. (1979): Communicating shapes by words and gestures. *Australian Journal of Psychology* 31, 137-147.
- Wandruszka, Mario (2001[1954]): Haltung und Gebärde der Romanen. In: Krüger, Reinhold (Hg.) (2001): *H. Urtel – L. Flachskampf – M. Wandruszka. Drei Studien zur Körpersprache der Romanen*, 105-179. Berlin: Weidler.
- Wang, Jenny Zhijie (1993): *The Geometry of Segmental Features in Beijing Mandarin*. Dissertation. Newark: University of Delaware. Zitiert nach Duanmu (2007: 346).
- Wängler, Hans-Heinrich (1983⁴): *Grundriß einer Phonetik des Deutschen. Mit einer allgemeinen Einführung in die Phonetik*. Marburg: Elwert.
- Watson, O. M. (1970): *Proxemic behavior: A cross-cultural study*. The Hague: Mouton.
- Webb, Jamer T. (1969): Subject speech rates as a function of interviewer behaviour. *Language and Speech* 21 (1), 54-67.
- Weiss, Benjamin (2008): *Sprechtempoabhängige Aussprachevariationen*. Dissertation. Humboldt-Universität zu Berlin. Publikation im Internet: <http://edoc.hu-berlin.de/dissertationen/weiss-benjamin-2008-05-28/PDF/weiss.pdf> ; Zugriff: 26.04.2011.
- Westbrook Eakins, Barbara / Eakins, Gene (1982): Sex differences in nonverbal communication. In: Samovar, Larry A. / Porter, Richard E. (Hg.): *Intercultural Communication: A reader*, 292-309. Belmont: Wadsworth Publishing Company. Zitiert nach Scollon/Wong Scollon (2002²: 301).
- White, Laurence / Mattys, Sven L. (2007): Calibrating rhythm: First language and second language studies. *Journal of Phonetics* 35, 501-532.
- Whiteside, Sandra P. (1996): Temporal-based acoustic-phonetic patterns in read speech: some evidence for speaker sex differences. *Journal of the International Phonetic Association* 26, 23-40.
- Wieneke, G. / Jansse, P. / Belderbos, H. (1987): The influence of speaking rate in the duration of jaw movements. *Journal of Phonetics* 15, 111-126.
- Wierzbicka, Anna (1999): *Emotions Across Languages and Cultures: Diversity and Universals*. Cambridge: University Press.
- Wierzbicka, Anna (2003²): *Cross-Cultural Pragmatics. The Semantics of Human Interaction*. Berlin / New York: Mouton de Gruyter.
- Wiget, Lukas / White, Laurence / Schuppler, Barbara / Grenon, Isabelle / Rauch, Olesya / Mattys, Sven L. (2010): How stable are acoustic metrics of contrastive speech rhythm? *Journal of the Acoustic Society of America* 127 (3), 1559-1569.
- Wood, Sidney (1973a): Speech Tempo. *Phonetics Laboratory Lund University Working Papers* 9, 99-147. Lund: Phonetics Laboratory.
- Wood, Sidney (1973b): What happens to vowels and consonants when we speak faster? *Phonetics Laboratory Lund University Working Papers* 9, 8-39. Lund: Phonetics Laboratory.
- Woodall, Gill W. / Burgoon, Judee K. (1983): Talking fast and changing attitudes: A Critique and Clarification. *Journal of Nonverbal Behavior* 8 (2), 126-142.
- Xu, Yi (1989): *Distribution of vowel information in prevocalic fricative noise*. Manuskript. University of Connecticut, Storrs. Zitiert nach Manuel in Hardcastle (1999: 381).
- Yanagawa, Mariko (2006): *Articulatory timing in first and second language: A cross-linguistic study*. Unveröffentlichte Dissertation. Yale University.

- Yang, Lichiung (2007): Duration, Pauses and the Temporal Structure of Mandarin Conversational Speech. *16th ICPHS (Saarbrücken)*, 1289-1292.
- Yen, Cha-jen (1992): *Kontrastive Untersuchung zur segmentalen Phonetik und Phonologie des Chinesischen und Deutschen*. Erlangen: Palm & Enke.
- Yuan, Jiahong / Liberman, Mark / Cieri, Christopher (2006): Towards an Integrated Understanding of Speaking Rate in Conversation. *Proceedings of Interspeech 2006*, 541-544.
- Zandpour, Fred / Sadry, Golnaz (1996): Communication in personal relationships in Iran: A comparative analysis. In: Gudykunst, William B. / Ting-Toomey, Stella / Nishida, Tsukasa: *Communication in personal relationships across cultures*, 174-196. Thousand Oaks: Sage.
- Zellner, Brigitte (1998): Fast and Slow Speech Rate: a Characterisation for French. *ICSLP, 5th International Conference on Spoken Language Processing (Sydney, Australia) 7*, 3159-3163.
- Zsiga, Elizabeth C. (2000): Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics* 28, 69-102.
- Zwaardemaker, H. / Eijkman, L. (1928): *Leerboek der Phonetiek: In-zonderheid met betrekking tot het Standaard-Nederlandsch*. Haarlem: Bohn.

Anhang

Tabelle 10.1: Rohdaten der Sprechtempomesswerte: Mittelwerte pro Sprecher

Spra- che	Ge- schl.	AG (Laute)	AG (Silben)	SG (Laute)	SG (Silben)	SG (Mor- pheme)	SG (Wörter)	Laute /Silbe	Morphe- me/Silbe	Silben /Wort	Morphe- me/Wort	%Pausen	%interne Pausen	s/art. Phrase	Silben/art. Phrase	Silben- dauer	Pausen- dauer	s/ Pause
BP	W	16,486	7,083	14,380	6,178	4,680	2,977	2,327	0,758	2,075	1,572	12,780	5,556	2,626	18,600	0,141	0,392	2,674
BP	W	16,437	6,956	14,611	6,183	4,963	3,108	2,363	0,803	1,990	1,597	11,111	8,333	4,202	29,231	0,144	0,569	4,553
BP	W	16,026	6,941	14,565	6,309	4,502	2,911	2,309	0,714	2,167	1,546	9,116	6,667	2,709	18,806	0,144	0,281	2,800
BP	W	16,179	6,947	14,260	6,122	4,547	2,847	2,329	0,743	2,150	1,597	11,866	12,903	2,447	17,000	0,144	0,340	2,526
BP	W	14,728	6,448	13,138	5,752	4,309	2,703	2,284	0,749	2,128	1,594	10,796	7,500	3,196	20,610	0,155	0,396	3,276
BP	W	16,475	7,188	15,614	6,812	5,644	3,136	2,292	0,829	2,172	1,800	5,225	0,000	3,984	28,636	0,139	0,242	4,382
BP	M	15,454	6,572	13,406	5,701	4,564	2,679	2,352	0,800	2,128	1,704	13,251	8,511	2,733	17,958	0,152	0,426	2,791
BP	M	17,126	7,174	15,480	6,485	4,837	3,296	2,387	0,746	1,967	1,467	9,607	0,000	3,668	26,313	0,139	0,416	3,912
BP	M	16,389	7,025	15,275	6,548	4,858	3,025	2,333	0,742	2,165	1,606	6,794	4,348	3,814	26,792	0,142	0,290	3,979
BP	M	14,948	6,388	12,720	5,436	3,963	2,571	2,340	0,729	2,115	1,542	14,907	10,909	1,702	10,875	0,157	0,304	1,733
BP	M	14,793	6,393	12,619	5,454	4,264	2,727	2,314	0,782	2,000	1,564	14,696	3,448	2,774	17,733	0,156	0,494	2,869
BP	M	16,382	6,907	14,352	6,051	4,814	2,947	2,372	0,796	2,053	1,634	12,393	5,556	2,050	14,158	0,145	0,306	2,164
CN	W	19,584	6,027	16,947	5,216	5,412	2,348	3,249	1,038	2,078	2,156	13,466	16,923	1,473	8,879	0,166	0,352	1,473
CN	W	19,451	6,132	17,645	5,563	5,563	2,440	3,172	1,000	2,119	2,119	9,285	13,514	2,146	13,158	0,163	0,411	2,146
CN	W	19,243	6,036	17,543	5,503	5,574	2,487	3,188	1,013	2,073	2,099	8,831	15,000	2,194	13,244	0,166	0,384	2,194
CN	W	18,740	5,981	16,739	5,342	5,266	2,347	3,133	0,986	2,153	2,123	10,679	22,000	1,843	11,020	0,167	0,345	1,843
CN	W	18,142	5,849	15,782	5,088	5,137	2,130	3,101	1,010	2,389	2,412	13,009	9,195	1,838	10,750	0,171	0,379	1,838
CN	W	17,122	5,324	15,528	4,828	4,865	2,334	3,216	1,008	1,977	1,992	9,313	6,250	1,986	10,571	0,188	0,311	1,986
CN	W	18,743	6,104	17,578	5,724	5,786	2,500	3,071	1,011	2,209	2,232	6,217	17,143	2,121	12,944	0,164	0,230	2,121
CN	M	18,343	5,841	15,936	5,075	5,157	2,382	3,140	1,016	2,061	2,094	13,122	33,333	1,588	9,275	0,171	0,309	1,588
CN	M	16,637	5,208	15,058	4,714	4,714	2,056	3,195	1,000	2,159	2,159	9,488	16,667	2,100	10,935	0,192	0,375	2,100
CN	M	17,908	5,932	16,455	5,451	5,440	2,395	3,019	0,998	2,206	2,202	8,114	20,690	2,703	16,033	0,169	0,342	2,703
CN	M	18,857	5,968	16,167	5,116	5,167	2,288	3,160	1,010	1,894	1,913	14,266	14,433	1,561	9,316	0,168	0,315	1,561
CN	M	17,998	5,873	17,998	5,873	5,879	2,505	3,064	1,001	2,140	2,142	12,585	20,000	2,127	12,494	0,170	0,478	2,127
CN	M	19,424	6,195	16,325	5,206	5,343	2,752	3,136	1,026	2,251	2,310	15,956	15,789	1,272	7,879	0,161	0,389	1,272
CN	M	18,288	6,053	16,198	5,361	5,413	2,346	3,022	1,010	2,193	2,215	11,432	15,254	1,407	8,517	0,165	0,252	1,407
DE	W	14,571	4,962	12,527	4,266	3,659	1,719	2,937	0,858	2,481	2,128	14,026	0,000	3,811	18,911	0,202	0,636	3,811
DE	W	15,832	5,256	13,911	4,618	3,884	2,292	3,012	0,841	2,015	1,695	12,135	13,043	3,242	17,042	0,190	0,467	3,242
DE	W	16,532	5,240	14,216	4,506	3,996	2,227	3,155	0,887	2,023	1,794	14,008	0,000	3,592	18,826	0,191	0,612	3,592
DE	W	15,597	5,163	13,614	4,507	4,141	2,280	3,021	0,919	1,977	1,817	12,714	0,000	3,478	17,958	0,194	0,529	3,478
DE	W	14,817	5,199	12,727	4,466	3,844	2,019	2,850	0,861	2,212	1,904	14,108	11,111	4,657	24,211	0,192	0,807	4,657
DE	W	14,896	5,097	13,434	4,597	3,809	2,212	2,923	0,829	2,078	1,722	9,811	10,000	3,980	20,286	0,196	0,455	3,980

DE	W	14,640	4,903	13,059	4,374	4,061	2,239	2,986	0,929	1,953	1,814	10,801	0,000	4,079	20,000	0,204	0,519	4,079
DE	M	16,086	5,376	13,478	4,504	3,376	2,119	2,992	0,750	2,126	1,594	16,215	1,961	2,357	12,673	0,186	0,465	2,357
DE	M	15,227	5,243	13,261	4,566	3,629	2,086	2,904	0,795	2,189	1,740	12,909	0,000	4,091	21,450	0,191	0,638	4,091
DE	M	15,094	5,412	13,627	4,886	3,995	2,308	2,789	0,818	2,117	1,731	9,723	0,000	5,137	27,800	0,185	0,593	5,137
DE	M	15,199	5,046	12,679	4,209	3,708	2,128	3,012	0,881	1,978	1,743	16,584	7,843	3,106	15,673	0,198	0,630	3,106
DE	M	14,991	5,196	12,658	4,388	3,725	2,252	2,885	0,849	1,949	1,654	15,561	0,000	3,821	19,857	0,192	0,739	3,821
DE	M	14,620	4,887	13,024	4,353	3,833	2,248	2,992	0,880	1,937	1,705	10,918	0,000	5,379	26,286	0,205	0,710	5,379
EN	W	15,143	5,304	13,381	4,687	3,766	2,707	2,855	0,804	1,731	1,391	11,637	4,878	3,497	18,548	0,189	0,472	3,497
EN	W	14,909	5,131	12,844	4,420	3,734	2,626	2,906	0,845	1,683	1,422	13,853	5,769	2,890	14,830	0,195	0,474	2,890
EN	W	15,284	5,275	13,340	4,604	3,802	2,732	2,898	0,826	1,685	1,391	12,719	5,000	3,662	19,317	0,190	0,547	3,662
EN	W	14,813	5,141	12,897	4,476	3,792	2,737	2,882	0,847	1,635	1,385	12,933	2,000	2,994	15,392	0,195	0,454	2,994
EN	W	15,348	5,376	13,282	4,652	3,899	2,935	2,855	0,838	1,585	1,329	13,462	0,000	2,794	15,023	0,186	0,445	2,794
EN	W	14,807	4,902	13,087	4,333	3,731	2,719	3,020	0,861	1,594	1,372	11,615	11,538	2,852	13,981	0,204	0,382	2,852
EN	M	14,853	5,136	13,148	4,546	3,872	2,790	2,892	0,852	1,629	1,388	11,479	8,889	3,255	16,717	0,195	0,431	3,255
EN	M	14,869	5,273	13,204	4,683	3,801	2,775	2,820	0,812	1,688	1,370	11,193	7,317	3,955	20,857	0,190	0,511	3,955
EN	M	15,687	5,383	13,917	4,775	3,894	2,880	2,914	0,815	1,658	1,352	11,284	15,094	2,852	15,352	0,186	0,370	2,852
EN	M	14,981	5,273	12,964	4,563	3,779	2,878	2,841	0,828	1,585	1,313	13,463	16,129	2,348	12,381	0,190	0,371	2,348
EN	M	15,701	5,303	13,714	4,632	3,850	2,789	2,961	0,831	1,660	1,380	12,652	1,695	2,813	14,917	0,189	0,414	2,813
EN	M	15,198	5,280	13,358	4,641	3,890	2,798	2,878	0,838	1,659	1,390	12,105	6,818	3,148	16,622	0,189	0,443	3,148
EP	W	11,888	5,170	11,078	4,818	3,864	2,422	2,299	0,802	1,989	1,596	6,817	37,500	4,019	20,778	0,193	0,331	4,019
EP	W	15,400	6,671	14,793	6,408	5,300	3,204	2,309	0,827	2,000	1,654	3,940	50,000	4,857	32,400	0,150	0,249	4,857
EP	W	13,376	5,504	12,282	5,054	4,751	2,780	2,430	0,940	1,818	1,709	8,176	9,091	3,028	16,667	0,182	0,294	3,028
EP	W	14,551	5,928	12,149	4,949	4,422	2,744	2,455	0,893	1,804	1,611	16,511	18,750	2,289	13,569	0,169	0,460	2,289
EP	W	14,544	6,242	13,328	5,720	5,053	2,962	2,330	0,883	1,931	1,706	8,360	15,385	3,002	18,741	0,160	0,284	3,002
EP	W	14,237	6,002	12,978	5,471	4,402	2,697	2,372	0,805	2,028	1,632	8,843	14,286	2,388	14,333	0,167	0,248	2,388
EP	W	12,479	5,017	11,969	4,812	4,060	2,316	2,488	0,844	2,078	1,753	4,087	0,000	6,379	32,000	0,199	0,340	6,379
EP	W	13,863	5,828	12,360	5,196	4,715	2,930	2,379	0,907	1,773	1,609	10,841	12,903	2,435	14,188	0,172	0,306	2,435
EP	M	14,148	6,182	12,739	5,567	4,645	2,874	2,288	0,834	1,937	1,616	9,956	31,429	3,037	18,778	0,162	0,345	3,037
EP	M	14,132	5,732	11,228	4,554	4,004	2,484	2,466	0,879	1,833	1,612	20,552	13,889	3,008	17,243	0,174	0,800	3,008
EP	M	16,108	6,815	14,495	6,133	4,971	3,159	2,364	0,811	1,941	1,574	10,012	20,000	1,761	12,000	0,147	0,216	1,761
EP	M	15,338	6,848	13,478	6,017	5,085	3,064	2,240	0,845	1,964	1,659	12,130	17,647	2,199	15,056	0,146	0,321	2,199
EP	M	15,984	6,416	13,972	5,609	5,224	2,979	2,491	0,931	1,883	1,754	12,585	35,135	2,572	16,500	0,156	0,380	2,572
EP	M	16,134	6,762	13,203	5,533	4,989	2,841	2,386	0,902	1,947	1,756	18,166	14,815	2,736	18,500	0,148	0,630	2,736
EP	M	13,659	5,730	12,110	5,080	4,568	2,685	2,384	0,899	1,892	1,702	11,343	25,000	2,618	15,000	0,175	0,343	2,618
EP	M	15,128	6,442	10,739	4,572	3,725	2,335	2,349	0,815	1,958	1,595	29,017	16,505	1,473	9,490	0,155	0,608	1,473

ES	W	17,045	7,333	15,220	6,548	4,914	3,293	2,324	0,750	1,989	1,492	10,707	0,000	4,736	34,733	0,136	0,609	4,736
ES	W	17,714	7,435	15,358	6,446	5,115	3,297	2,383	0,794	1,955	1,551	13,304	20,930	3,436	25,545	0,135	0,540	3,436
ES	W	16,538	7,060	13,421	5,729	4,191	2,681	2,343	0,731	2,137	1,563	18,848	25,532	2,352	16,604	0,142	0,558	2,352
ES	W	16,369	6,937	12,928	5,478	4,254	2,778	2,360	0,776	1,972	1,531	21,025	15,686	2,728	18,923	0,144	0,740	2,728
ES	W	17,042	7,281	14,771	6,310	4,904	3,166	2,341	0,777	1,993	1,549	13,329	35,897	1,490	10,848	0,137	0,232	1,490
ES	W	17,212	7,457	14,300	6,195	4,615	2,918	2,308	0,745	2,123	1,581	16,918	10,870	2,907	21,681	0,134	0,605	2,907
ES	M	16,281	6,881	12,823	5,420	4,141	2,703	2,366	0,764	2,005	1,532	21,237	10,638	2,465	16,958	0,145	0,679	2,465
ES	M	17,432	7,261	15,227	6,343	4,768	3,284	2,401	0,752	1,932	1,452	12,647	13,333	4,855	35,250	0,138	0,750	4,855
ES	M	17,536	7,373	13,568	5,704	4,542	2,856	2,378	0,796	1,997	1,591	22,631	5,000	2,517	18,561	0,136	0,755	2,517
ES	M	17,158	7,208	14,528	6,103	4,645	3,213	2,381	0,761	1,899	1,445	15,332	7,692	2,240	16,143	0,139	0,437	2,240
ES	M	16,542	7,007	13,485	5,712	4,170	2,838	2,361	0,730	2,013	1,470	18,482	17,742	2,098	14,698	0,143	0,483	2,098
ES	M	17,928	7,746	14,985	6,474	4,980	3,373	2,315	0,769	1,920	1,476	16,414	5,882	2,556	19,800	0,129	0,517	2,556

Tabelle 10.2: Einfaktorielle Varianzanalyse: F-Werte (Variablen: Tempomesswerte, Faktor: Sprache)

Variable		Quadrat summe	df	Mittel der Quadrate	F	Signifikanz
AG (Silben/s)	Zwischen den Gruppen	43,495	5	8,699	78,203	0
	Innerhalb der Gruppen	8,12	73	0,111		
	Gesamt	51,615	78			
SG (Laute/s)	Zwischen den Gruppen	134,81	5	26,962	34,963	0
	Innerhalb der Gruppen	56,294	73	0,771		
	Gesamt	191,104	78			
SG (Morpheme/s)	Zwischen den Gruppen	22,037	5	4,407	36,365	0
	Innerhalb der Gruppen	8,847	73	0,121		
	Gesamt	30,884	78			
SG (Wörter/s)	Zwischen den Gruppen	7,132	5	1,426	32,917	0
	Innerhalb der Gruppen	3,163	73	0,043		
	Gesamt	10,295	78			
Pausenanteil	Zwischen den Gruppen	274,694	5	54,939	3,821	0,004
	Innerhalb der Gruppen	1049,474	73	14,376		
	Gesamt	1324,168	78			
Anteil interner Pausen	Zwischen den Gruppen	3281,176	5	656,235	10,3	0
	Innerhalb der Gruppen	4650,893	73	63,711		
	Gesamt	7932,069	78			

Tabelle 10.3: Einfaktorielle Varianzanalyse: Post-Hoc-Vergleiche nach Scheffé (Variablen: Tempomesswerte, Faktor: Sprache)

Tempo- parameter	Spra- che1	Spra- che2	Mittlere Differenz	Standard- fehler	Signifi- kanz	95%- Konfidenzintervall	
						Unter- grenze	Ober- grenze
AG (Silben/s)	BP	CN	,94081644*	0,13120595	0	0,4920204	1,389613
		DE	1,68297037*	0,13351474	0	1,226277	2,139664
		EN	1,60388670*	0,13615885	0	1,138149	2,069624
		EP	,75479407*	0,12736494	0	0,3191364	1,190452
		ES	-0,41293939	0,13615885	0,116	-0,878677	0,052798
	CN	BP	-,94081644*	0,13120595	0	-1,389613	-0,492020
		DE	,74215392*	0,12845993	0	0,3027508	1,181557
		EN	,66307025*	0,13120595	0	0,2142742	1,111866
		EP	-0,18602237	0,12205572	0,801	-0,60352	0,231475
		ES	-1,35375584*	0,13120595	0	-1,802552	-0,90496
	DE	BP	-1,68297037*	0,13351474	0	-2,139664	-1,226277
		CN	-,74215392*	0,12845993	0	-1,181557	-0,302751
		EN	-0,07908367	0,13351474	0,996	-0,535777	0,37761
		EP	-,92817629*	0,12453426	0	-1,354152	-0,502201
		ES	-2,09590976*	0,13351474	0	-2,552603	-1,639216
	EN	BP	-1,60388670*	0,13615885	0	-2,069624	-1,138149
		CN	-,66307025*	0,13120595	0	-1,111866	-0,214274
		DE	0,07908367	0,13351474	0,996	-0,37761	0,535777
		EP	-,84909262*	0,12736494	0	-1,284750	-0,413435
		ES	-2,01682609*	0,13615885	0	-2,482564	-1,551088
	EP	BP	-,75479407*	0,12736494	0	-1,190452	-0,319136
		CN	0,18602237	0,12205572	0,801	-0,231475	0,60352
		DE	,92817629*	0,12453426	0	0,5022011	1,354152
		EN	,84909262*	0,12736494	0	0,4134349	1,284750
		ES	-1,16773347*	0,12736494	0	-1,603391	-0,732076

	ES	BP	0,41293939	0,13615885	0,116	-0,052798	0,878677
		CN	1,35375584*	0,13120595	0	0,9049598	1,802552
		DE	2,09590976*	0,13351474	0	1,6392164	2,552601
		EN	2,01682609*	0,13615885	0	1,5510884	2,482564
		EP	1,16773347*	0,12736494	0	0,7320758	1,603391
SG (Laute/s)	BP	CN	-2,36235844*	0,34546265	0	-3,54403	-1,180687
		DE	0,95433561	0,35154164	0,209	-0,248129	2,156800
		EN	0,94022143	0,35850354	0,243	-0,286057	2,166499
		EP	1,52030721*	0,33534935	0,002	0,3732292	2,667385
		ES	-0,01610009	0,35850354	1	-1,242378	1,210178
	CN	BP	2,36235844*	0,34546265	0	1,1806874	3,54403
		DE	3,31669405*	0,33823244	0	2,1597543	4,473634
		EN	3,30257987*	0,34546265	0	2,1209088	4,484251
		EP	3,88266565*	0,32137026	0	2,7834037	4,981928
		ES	2,34625835*	0,34546265	0	1,1645873	3,527929
	DE	BP	-0,95433561	0,35154164	0,209	-2,1568	0,248129
		CN	-3,31669405*	0,33823244	0	-4,473634	-2,159754
		EN	-0,01411418	0,35154164	1	-1,216579	1,188350
		EP	0,5659716	0,32789621	0,703	-0,555613	1,687556
		ES	-0,9704357	0,35154164	0,193	-2,1729	0,232029
	EN	BP	-0,94022143	0,35850354	0,243	-2,166499	0,286057
		CN	-3,30257987*	0,34546265	0	-4,484251	-2,120909
		DE	0,01411418	0,35154164	1	-1,18835	1,216579
		EP	0,58008578	0,33534935	0,701	-0,566992	1,727164
		ES	-0,95632152	0,35850354	0,226	-2,1826	0,269956
	EP	BP	-1,52030721*	0,33534935	0,002	-2,667385	-0,373229
		CN	-3,88266565*	0,32137026	0	-4,981928	-2,783404
		DE	-0,5659716	0,32789621	0,703	-1,687556	0,555613
		EN	-0,58008578	0,33534935	0,701	-1,727164	0,566992
		ES	-1,53640730*	0,33534935	0,002	-2,683485	-0,389329
	ES	BP	0,01610009	0,35850354	1	-1,210178	1,242378
		CN	-2,34625835*	0,34546265	0	-3,527929	-1,164587
		DE	0,9704357	0,35154164	0,193	-0,232029	2,1729
		EN	0,95632152	0,35850354	0,226	-0,269957	2,1826
		EP	1,53640730*	0,33534935	0,002	0,3893293	2,683485
SG (Morpheme/s)	BP	CN	-,67469087*	0,13695549	0,001	-1,143154	-0,206228
		DE	,84185570*	0,13936545	0	0,3651497	1,318562
		EN	,84445213*	0,14212542	0	0,3583055	1,3305987
		EP	0,05088659	0,13294616	1	-0,403862	0,505635
		ES	0,05877018	0,14212542	0,999	-0,427376	0,544917
	CN	BP	,67469087*	0,13695549	0,001	0,2062283	1,143154
		DE	1,51654657*	0,13408914	0	1,0578885	1,975205
		EN	1,51914300*	0,13695549	0	1,0506804	1,987606
		EP	,72557746*	0,12740428	0	0,2897852	1,16137
		ES	,73346105*	0,13695549	0	0,2649984	1,201924
	DE	BP	-,84185570*	0,13936545	0	-1,318562	-0,36515
		CN	-1,51654657*	0,13408914	0	-1,975205	-1,057889
		EN	0,00259643	0,13936545	1	-0,47411	0,479302
		EP	-,79096911*	0,12999143	0	-1,235611	-0,346327
		ES	-,78308552*	0,13936545	0	-1,259792	-0,306385
	EN	BP	-,84445213*	0,14212542	0	-1,330599	-0,358306
		CN	-1,51914300*	0,13695549	0	-1,987606	-1,050684
		DE	-0,00259643	0,13936545	1	-0,479302	0,47411

		EP	-,79356554*	0,13294616	0	-1,248314	-0,338817
		ES	-,78568195*	0,14212542	0	-1,271827	-0,299535
	EP	BP	-0,05088659	0,13294616	1	-0,505635	0,403862
		CN	-,72557746*	0,12740428	0	-1,16137	-0,289785
		DE	,79096911*	0,12999143	0	0,3463274	1,235611
		EN	,79356554*	0,13294616	0	0,338817	1,248314
		ES	0,00788359	0,13294616	1	-0,446865	0,462632
	ES	BP	-0,05877018	0,14212542	0,999	-0,544917	0,427376
		CN	-,73346105*	0,13695549	0	-1,201924	-0,264998
		DE	,78308552*	0,13936545	0	0,3063795	1,259792
		EN	,78568195*	0,14212542	0	0,2995353	1,271829
		EP	-0,00788359	0,13294616	1	-0,462632	0,446865
SG (Wörter/s)	BP	CN	,53114128*	0,08189135	0	0,2510281	0,811254
		DE	,74672155*	0,08333236	0	0,4616793	1,031764
		EN	0,12980417	0,08498267	0,8	-0,160883	0,420491
		EP	0,13075289	0,07949401	0,745	-0,14116	0,402666
		ES	-0,1229659	0,08498267	0,834	-0,413653	0,167721
	CN	BP	-,53114128*	0,08189135	0	-0,811254	-0,251028
		DE	0,21558027	0,08017744	0,218	-0,05867	0,489831
		EN	-,40133711*	0,08189135	0,001	-0,68145	-0,121224
		EP	-,40038839*	0,07618029	0	-0,660967	-0,13981
		ES	-,65410718*	0,08189135	0	-0,93422	-0,373994
	DE	BP	-,74672155*	0,08333236	0	-1,031764	-0,461679
		CN	-0,21558027	0,08017744	0,218	-0,489831	0,05867
		EN	-,61691738*	0,08333236	0	-0,90196	-0,331875
		EP	-,61596866*	0,07772726	0	-0,881838	-0,350099
		ES	-,86968745*	0,08333236	0	-1,15473	-0,584645
	EN	BP	-0,12980417	0,08498267	0,8	-0,420491	0,160883
		CN	,40133711*	0,08189135	0,001	0,1212239	0,68145
		DE	,61691738*	0,08333236	0	0,3318752	0,90196
		EP	0,00094873	0,07949401	1	-0,270964	0,272862
		ES	-0,25277007	0,08498267	0,13	-0,543457	0,037917
	EP	BP	-0,13075289	0,07949401	0,745	-0,402666	0,14116
		CN	,40038839*	0,07618029	0	0,1398102	0,660967
		DE	,61596866*	0,07772726	0	0,350099	0,881838
		EN	-0,00094873	0,07949401	1	-0,272862	0,270964
		ES	-0,2537188	0,07949401	0,083	-0,525632	0,018194
	ES	BP	0,1229659	0,08498267	0,834	-0,167721	0,413653
		CN	,65410718*	0,08189135	0	0,373994	0,93422
		DE	,86968745*	0,08333236	0	0,5846452	1,15473
		EN	0,25277007	0,08498267	0,13	-0,037917	0,543457
		EP	0,2537188	0,07949401	0,083	-0,018194	0,525632
Pausenanteil	BP	CN	-0,08081081	1,49161377	1	-5,182945	5,021323
		DE	-1,99431446	1,51786119	0,884	-7,186229	3,1976
		EN	-1,32118763	1,54792076	0,981	-6,615923	3,973547
		EP	-0,91317874	1,44794729	0,995	-5,86595	4,039592
		ES	-5,69435942*	1,54792076	0,027	-10,98909	-0,399625
	CN	BP	0,08081081	1,49161377	1	-5,021323	5,182945
		DE	-1,91350366	1,46039568	0,885	-6,908855	3,081848
		EN	-1,24037683	1,49161377	0,983	-6,342511	3,861757
		EP	-0,83236793	1,38758939	0,996	-5,578682	3,913946
		ES	-5,61354861*	1,49161377	0,022	-10,71568	-0,511415
	DE	BP	1,99431446	1,51786119	0,884	-3,1976	7,186229

		CN	1,91350366	1,46039568	0,885	-3,081848	6,908855
		EN	0,67312683	1,51786119	0,999	-4,518788	5,865042
		EP	1,08113573	1,41576667	0,988	-3,76156	5,923831
		ES	-3,70004496	1,51786119	0,323	-8,89196	1,49187
	EN	BP	1,32118763	1,54792076	0,981	-3,973547	6,615923
		CN	1,24037683	1,49161377	0,983	-3,861757	6,342511
		DE	-0,67312683	1,51786119	0,999	-5,865042	4,518788
		EP	0,40800889	1,44794729	1	-4,544762	5,36078
		ES	-4,37317179	1,54792076	0,172	-9,667907	0,921563
	EP	BP	0,91317874	1,44794729	0,995	-4,039592	5,86595
		CN	0,83236793	1,38758939	0,996	-3,913946	5,578682
		DE	-1,08113573	1,41576667	0,988	-5,923831	3,76156
		EN	-0,40800889	1,44794729	1	-5,36078	4,544762
		ES	-4,78118068	1,44794729	0,065	-9,733952	0,171590
	ES	BP	5,69435942*	1,54792076	0,027	0,3996246	10,98909
		CN	5,61354861*	1,49161377	0,022	0,5114145	10,71568
		DE	3,70004496	1,51786119	0,323	-1,491868	8,89196
		EN	4,37317179	1,54792076	0,172	-0,921563	9,667907
		EP	4,78118068	1,44794729	0,065	-0,17159	9,733952
Interne Pausen	BP	CN	-10,7266197	3,14006607	0,051	-21,46736	0,014122
		DE	2,76275675	3,19532074	0,98	-8,166986	13,6925
		EN	-0,94983367	3,25860056	1	-12,09603	10,19636
		EP	-14,62668108*	3,04814171	0,001	-25,05299	-4,200371
		ES	-7,95612353	3,25860056	0,321	-19,10232	3,190071
	CN	BP	10,7266197	3,14006607	0,051	-0,014122	21,46736
		DE	13,48937649*	3,0743474	0,004	2,9734285	24,00533
		EN	9,77678607	3,14006607	0,098	-0,963956	20,51753
		EP	-3,90006134	2,92107946	0,876	-13,89175	6,091627
		ES	2,77049621	3,14006607	0,978	-7,970246	13,51124
	DE	BP	-2,76275675	3,19532074	0,98	-13,6925	8,166986
		CN	-13,48937649*	3,0743474	0,004	-24,00532	-2,973429
		EN	-3,71259042	3,19532074	0,928	-14,64233	7,217152
		EP	-17,38943783*	2,98039677	0	-27,58402	-7,194852
		ES	-10,7188803	3,19532074	0,058	-21,64862	0,210863
	EN	BP	0,94983367	3,25860056	1	-10,19636	12,09603
		CN	-9,77678607	3,14006607	0,098	-20,51753	0,963956
		DE	3,71259042	3,19532074	0,928	-7,217153	14,64233
		EP	-13,67684741*	3,04814171	0,003	-24,10316	-3,250537
		ES	-7,00628986	3,25860056	0,47	-18,15248	4,139905
	EP	BP	14,62668108*	3,04814171	0,001	4,2003708	25,05299
		CN	3,90006134	2,92107946	0,876	-6,09163	13,89175
		DE	17,38943783*	2,98039677	0	7,1948523	27,58402
		EN	13,67684741*	3,04814171	0,003	3,2505372	24,10316
		ES	6,67055755	3,04814171	0,449	-3,755753	17,09687
	ES	BP	7,95612353	3,25860056	0,321	-3,190071	19,10232
		CN	-2,77049621	3,14006607	0,978	-13,51124	7,970246
		DE	10,7188803	3,19532074	0,058	-0,210863	21,64862
		EN	7,00628986	3,25860056	0,47	-4,139905	18,15248
		EP	-6,67055755	3,04814171	0,449	-17,09687	3,755753

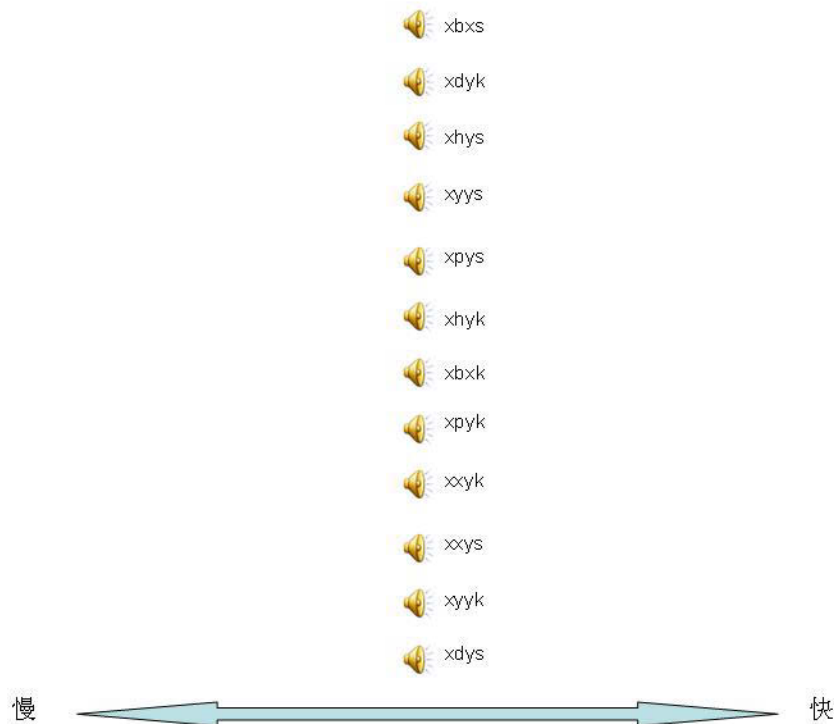


Abbildung 10.1: Grafische Oberfläche des Wahrnehmungsexperiments (5.5) mit chinesischen Teilnehmern

Tabelle 10.4: Bewertung des Sprechtempos der Stimuli durch deutsche und chinesische Muttersprachler

Mut- ter- spr.	Aufnahme											
	BPs	DEk	CNs	ENs	PTs	CNk	BPk	PTk	ESk	ESs	ENk	DEs
DE	2	-4	-1	-3	3	2	1	-1	-3	2	-3	-2
DE	-2	-2	-1	1	0	-1	-1	-2	1,5	2,5	-3,5	-3
DE	-1	-2	-1	0	-1	-1,5	-1	-2	0,5	-1	-3	1
DE	-0,5	-3,5	1	4,5	-1,5	-1	-3	-5	1,5	2,5	-2	3,5
DE	1,5	-4	0,5	-0,5	2	-2,5	-3	1	2	3	-1	0
DE	-1	-2,5	-1,5	-1	-2	-1,5	-2,5	-3	0	0,5	-3,5	-2
DE	-1,5	-1	-1,5	2,5	2	-1	-1,5	-2	-1	2	-1	-1
DE	-1,5	-3	0	0	0	-3	0	0	3	4	-3	-1,5
DE	-0,5	-1,5	0,5	1,5	-1,5	-3	-2,5	-1	-4	1,5	-3,5	-2
DE	-3	-1,5	1,5	1	-1,5	1	-1	-2	2	2	-3	1
DE	0	-3	2	1,5	-2	-3	-1	-2	2	1	-1	-2
DE	-4,5	0	1,5	3	1	-1	-2	-3	-2	2	-1,5	1
DE	-1	-2	2	1	0	-3	-4	2	4	0,5	-2	2
DE	-1,5	-1	-0,5	0	-0,5	-2,5	-1	-1,5	-0,5	0	-3	0
DE	0,5	-1	1,5	1	-1	-1	1	1,5	0,5	0,5	-1	-1
DE	3,5	-3,5	2,5	-1,5	3	-1	-0,5	0	4	3	-4	-4
DE	1,5	-1	1,5	1	-1	-1,5	-1	0	-1	0,5	-2,5	-2
DE	2,5	-2	2	-1	0	-3	1	2,5	3,5	1,5	-1	1
DE	1	-2	2	2,5	1,5	-3	-3,5	-1	-1,5	4	-2	0,5
DE	1	-5	-1	-1,5	-3,5	-5,5	-4	-2	1,5	3,5	-2	2,5

DE	2	-2	-1	1	-0,5	-3	-1,5	-1	0,5	2,5	-3	0
DE	2	-2	-3,5	1,5	-1	-2,5	-2	-1	2	1	-4	-1
DE	0	-1	0,5	2	2	-0,5	0	1	0	3	-0,5	0
DE	2	0	3,5	0,5	-0,5	-2	-2	-2,5	-1	0,5	0	0
DE	-0,5	-2	1,5	1	2	-3	-2	1,5	4	5	0,5	1,5
DE	-0,5	-2	1,5	1	-1	-1	-1	-1	2	1	-2	1,5
DE	4	-3,5	3,5	1	3	-1,5	2,5	4	4	4,5	-3	1
DE	-1,5	-4	1	-2	1	-3	-2	2	3	4	-4,5	-1
DE	2	-0,5	0,5	1	0	-1,5	0	-1	-1,5	0,5	-2	-1
DE	0	-3	2,5	1,5	0	-1	0	2	3	2,5	-1	-2
DE	-1	0	-1	1	0,5	-1,5	-0,5	0	0	2,5	0	1,5
DE	3,5	-4	2	-2,5	0	-1,5	1	0	3	1	-2	1
DE	1,5	-1	2,5	1,5	2	-1	3	2	2,5	1,5	-2	-0,5
DE	-1,5	-0,5	1,5	-1	1,5	-1	1,5	2	1	2,5	-0,5	1,5
DE	3	-2	1	-3	2	-0,5	1,5	0	-0,5	1	-3	-3,5
DE	4	-4	4	-1,5	3	2	-2	3	4	4,5	-4	2
DE	2	-1,5	2,5	2,5	0	-2	0	2	2,5	3,5	-2	1
DE	-1,5	-2,5	3,5	-0,5	2	-2	-1,5	-1	3,5	2,5	-0,5	-1
DE	4	-4	4,5	3,5	2	-1,5	-2,5	3,5	0,5	2,5	-3	0,5
DE	2,5	-5	-1,5	-4,5	2	2	-1	2,5	-1	4	-4,5	-4
DE	-1,5	-2,5	1	0,5	-1	-2	-1	0	0	1	0	-1
DE	0,5	-2	1,5	0	1	-2,5	-1,5	-4	1	3	-3,5	-1,5
DE	0,5	-2	-1	-2	-0,5	-3	-2	-1	0	1	-3	-3
DE	0	-3	0	0	0	-3	-1,5	0	-2	0	-3	0,5
DE	0	0	2	0	2	1	-3	-2	-1	-2	0	1
DE	1	-2	4	4	1	-4	-4	3	-5	3	-5	0
DE	1	-2	-0,5	1	1	-2	-0,5	-1	1	1	-3	-2
DE	4	-5	4	4,5	4,5	2	-2	-5,5	-2	4,5	-5	4,5
DE	1	-3	-1	-0,5	0	-3,5	-3	-1,5	-2	0,5	-1,5	-0,5
DE	0,5	-3	1	-2	-2	-4	-1	0	-4	-1	-5	-1
DE	1,5	0,5	-2,5	-2	-2	1,5	1,5	1	-1	-2,5	-1	-2
DE	-1	-1,5	-2	1	-1	0,5	-0,5	1	2,5	1,5	-2	-1
DE	-1	-3	0	1	0	0	-4	-1	-1	3	-2	1
DE	0,5	0	2,5	2	2	-2,5	-2	-3	-1,5	2,5	-1,5	2,5
DE	0	-3	2	3	-2	-6	-1	1	-4	4	-5	2
DE	-3,5	-3	0	0,5	-0,5	-2	-1,5	1	0	2	-2	0,5
DE	0	-1	1	1,5	0	-1,5	0	-2,5	2,5	3,5	-3	2
DE	-4	0	-1	2	1,5	-3	-3	-5	-2	3	-2	1
DE	-2,5	2	3	3	1	1,5	-2	0	0	3	0	0
DE	1	-3	2	1	-1,5	-3	0	1	-1	2,5	-1,5	1
DE	3	-2	-1	2	-0,5	-3	-1	-4	-1,5	2,5	-0,5	1
DE	1	1	2	2	1,5	-0,5	1	-2	-1	1,5	-1	1
DE	-3	-3	-2	-2,5	-3	-2	-2	1	0	1,5	-2	-2,5
DE	1,5	1	3	1,2	1	1,5	1	-4	0	1	1,5	0
DE	-2	-3	3,5	1	0	-1,5	3	1,5	2	2	-2,5	-1
DE	-1,5	-2	-1	0	-1,5	0	1	0	-0,5	2	-2	1
DE	1,5	0	2	1,5	1	0	1	1	0	1,5	-0,5	1
DE	1,5	-1,5	1,5	2	2	-1,5	1	-1	-1	1	-2	-1,5
DE	-2,5	-3	-1	0,5	0	-2	-1	-1	0	0,5	-0,5	-1
DE	2	-3	-1	1	-1	-4	-4	-3	1	2	-3	1
DE	2	-1,5	2,5	1,5	1	-2,5	-2,5	2	2,5	4	-3,5	1
DE	3	-3,5	3	2	-2	-4	-3	-3	3	4	0	2
DE	0	0	0	-2,5	-1	-3,5	-2	-1	0	1	-3	0

DE	-4	2	1	-1,5	-1	1	-1,5	1	4,5	4	-1	0,5
DE	-5	-4	-3	-3	-3	-2,5	-2,5	-5	-1	3	1	2
DE	0	-1	1	-0,5	-1	-2	-1	-1	-0,5	1	-2	-2
CN	3	-1	1	3	-3	-5	3	1	-3	-1	-5	1
CN	-5	-3	-1	0	-2	-6	1	-4	4	3	1	2
CN	1	-3	-2	0	-1	-4	-1	3	2	1	-2	-3
CN	3	0	0	0	0	-4	-3	0	0	0	-2	-1
CN	-2	2	-5	0	-4	-6	-1	-1	5	4	-3	3
CN	-3	2	-5	4	3	-6	-1	2	-4	6	-2	4
CN	1	-1	-2	3	-2	-4	-3	1	2	2	-4	2
CN	-3	-3	1	0	0	-5	-5	1	4	4	-1	4
CN	2	-5	-1	-4	-3	-2	3	1	5	4	-6	0
CN	2	0	-1	-2	4	-5	1	3	6	5	-4	0
CN	-3	2	-1	0	3	-5	-3	-2	1	5	-4	4
CN	0	1	1	3	-3	-6	-4	2	5	4	3	0
CN	2	4	-1	1	5	-4	-3	6	-2	7	0	3
CN	-3	2	-2	1	0	-4	-1	1	3	2	0	1
CN	1,5	1	-2,5	2,5	2	-3	1	0	3	3	-2	1
CN	-1	0	1	2	2	-3	2	2	4	3	2	2
CN	2	-0,5	-3	2	-1	-3,5	0,5	1	2,5	2,5	-3	-0,5
CN	0,5	-1	0,5	2	0	-2,5	2,5	2	3	4	1,5	2
CN	-1	0	-2	0	0	-3	-2	1	3	2	-3	1
CN	-4	-1	-3	2	0	-5	-2	-1	3	4	-1	1
CN	0	1	-3	-4	-2	-5	4	4	2	5	-1	3
CN	-3	3,5	-3	3,5	3	-3,5	-3,5	5	6	6	-4	3
CN	1,5	-2	-3	1,5	-3	-5	3	4	5	-0,5	-4,5	0
CN	0,5	1	-3,5	1,5	-3	-4	2,5	-1	4	4	-0,5	0
CN	1	-1	0	0	2	-3	-1	3	3,5	3,5	0	1
CN	-1	2	-1	0	-3,5	-5	-1	-2,5	1	3	-4	-1,5
CN	-4	2	-3	3	-1	-5	-4	0	4	3	-2	1
CN	1,5	-2,5	-1	-2	-2	-5	-1	0	-3,5	2,5	-1	-1
CN	-3	2	-2	1	-1	-5	-4	-4	4	3	-6	0
CN	4	-4	-1	0	-2	-6	-3	2	5	3	-5	1
CN	3	-3	-5	2	1	-6	-2	-4	5	4	-1	0
CN	0	-1	-2	2	1	-4	-3	0	3	4	-3,5	2
CN	-2	0	-2	0	-3,5	-5,5	-3,5	2	-5	3,5	-5	5
CN	-2	-3	-4	0,5	-1,5	-5	1,5	2,5	4,5	3,5	-4,5	-0,5
CN	3	0	-2	2	-5	-6	1	-1	5	4	-4	-3
CN	1	-1	-5	-3	-4	-6	2	-1	5	4	0	3
CN	0	-1,5	-3,5	-1	-2	-4	-0,5	-2,5	1,5	1	-3	0,5
CN	-1	1	-4,5	0	0	-6	-2	-1	3,5	2,5	-3	0,5
CN	-4,5	-0,5	-1,5	2	0	-5,5	-3	1	4	2,5	1,5	-2
CN	2	-1	-3	1	-0,5	-4	0	3	3,5	4	-1,5	2
CN	-2	1	-3	3	-1,5	-4	-2	2,5	4	4	0,5	2
CN	-3	2	-4,5	4	-4	-5	1	4	1,5	4	-4	-2
CN	1	1	-1	0	-1	-5	-2	-3	4	3	-4	2
CN	2	2	0,5	3	-4	-4,5	1	0,5	4,5	1,5	-1	1
CN	2,5	-2	-3	1	0	-3,5	-2	1	4	3	-3	-1
CN	1	1	-1	0	-1	-5	-2	-3	4	3	-4	2
CN	4	2	-4	3	-5	-6	-4	-1	1	0	-2	-3
CN	2	1	-2	1	-3	-5	-1	-3	4	3	-4	0
CN	-2	1	-4	-1	0	-5	-3	0	4	3	-1	2
CN	2	-3	-4	2	1	-5,5	-4	2	4	3	-5	-3

CN	0	-3	0	2	1	-5,5	-4	-2	5	3	1	4
CN	1	2,5	-1	3	0,5	-1,5	-0,5	-2	1	2	-2,5	0
CN	-4	-1,5	-3	2	-1,5	-5,5	-3	-1,5	4	0	-5	0
CN	2	-3,5	-1	-2	-5	-6	0,5	-2,5	3	1,5	-4	0
CN	3	-2	-4	0	-3,5	-6	1,5	2,5	4,5	4	-1	0,5
CN	0,5	0	-0,5	0	-2	-2	0	-3	2	0	1	1,5
CN	0,5	-3	-4	-1	-1	-5,5	-2	-2	4	2	-5,5	-1
CN	1	0	-3	-1	-1,5	-4	2	1,5	3,5	3	-2	2,5
CN	-2,5	-3	-3	0,5	-1	-3,5	-1,5	0	1	0,5	-0,5	0,5
CN	-2	0	-4,5	3	0	-6	-2	-3	1,5	4	-3	3
CN	1,5	2	-3	0	3	-4	-1	0	3,5	4	-2	1
CN	3	2	-5	2,5	-3	-6	-1,5	1	5	4	-4	0
CN	4	-3	-4	1	-1	-6	0	1	3	2	-5	-2
CN	0	-1,5	-3,5	2,5	-1	-4	1	0,5	3	4	-2,5	2
CN	2,5	2	-2	1	0,5	-4	0	3,5	4	3,5	-3,5	1,5
CN	-3	4	1	3	-4	-5	-1	-2	2	2	0	-1
CN	-1	3,5	1,5	6	2	-4,5	-3	3,5	-6	-3	3	4,5
CN	2	-2	-2	-4,5	0	-6	4	1	5	5	-3	-3
CN	-4	-2	-4	2	-1	-5,5	2	-4	5	5	-1,5	5
CN	-1,5	1	0	1	2	-4,5	-3,5	3	4,5	4,5	-3	-2,5
CN	2,5	-4	1	-5,5	3	-5,5	4	-5,5	4	4	-5,5	4
CN	1,5	0,5	-3,5	-3	-2	-4	0	2	2	2	-3,5	-1
CN	1	0	0,5	0	-0,5	-2,5	1,5	0,5	3	2,5	-1,5	0
CN	0,5	-0,5	-1,5	-1	-0,5	-3,5	1	-0,5	3	2	0	1,5
CN	3	-2,5	-5	1	1	-6	-0,5	-4	5	3	-4	5

Tabelle 10.5: Lineares Regressionsmodell zur Erklärung der gemittelten wSG deutscher Muttersprachler

Modell		R	R ²	Korr. R ²	Standardfehler
		0,852	0,725	0,698	0,678633
ANOVA	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Regression	12,151	1	12,151	26,383	<0,001
Nicht standardisierte Residuen	4,605	10	0,461		
Gesamt	16,756	11			
KOEFFIZIENTEN	Regressionskoeffizient	Standardfehler	Beta	T	Signifikanz
Silben/s	0,718	0,14	0,852	5,136	<0,001

Tabelle 10.6: Lineares Regressionmodell zur Erklärung der gemittelten wSG chinesischer Muttersprachler

Modell		R	R ²	Korr. R ²	Standardfehler
		0,751	0,564	0,4	1,704943
ANOVA	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Regression	30,029	3	10,01	4,334	0,072
Nicht standardisierte Residuen	23,255	8	2,907		
Gesamt	53,283	11			
KOEFFIZIENTEN	Regressionskoeffizient	Standardfehler	Beta	T	Signifikanz
Laute/s	1,524	0,767	0,881	1,987	0,082
Silben/s	-2,01	1,731	-1,337	-1,161	0,279
Silbenkomplexität	-4,115	2,554	-1,711	-1,611	0,146

Tabelle 10.7: Rohdaten der Häufigkeit der Verwendung nonverbaler Kommunikationsmittel; fehlende Werte nicht erhoben

Sprache	Geschlecht	Gesten/min	Mimik/min	Bewegungen/min	Register
BP	M	9,28	0,64	0,64	formell
BP	W	3,34	10,9	14,3	formell
BP	W	7,06			formell
BP	M	6,96			formell
BP	M	1,34			formell
BP	W	21,95	9,85	12,25	semiformell
BP	M	18,45	13,3	3,3	semiinformell
BP	W	17,57	4,6	3,1	semiinformell
BP	M	4,01	4,6	7,3	semiformell
BP	M	11	9,27	10,99	formell
BP	M	10,9	5,5	8	semiformell
CN	M	0	0	5,5	semiformell
CN	M	0,33			semiinformell
CN	W	0			formell
CN	W	3,05	0	29,4	informell
CN	W	0	5,28	25,93	formell
CN	W	0			semiformell
CN	W	0			semiformell
CN	M	0	5,4	11,8	semiformell
CN	W	8,42			informell
CN	M	0	4	23	formell
CN	M	0			formell
CN	W	0			semiformell
CN	W	0			semiformell
CN	W	0			semiformell
CN	W	0			semiformell
CN	W	0			semiformell
CN	M	0			semiformell
CN	W	0	11,66	13,03	formell
CN	M	0,78			semiinformell
CN	W	1,54	3,8	16,7	semiformell
CN	W	0	3,2	18,9	semiformell
CN	M	0	7,5	19,7	semiformell
CN	M	0	1,5	21,35	formell
CN	W	0	10,7	27,9	semiformell
CN	M	0	4,8	33,6	formell
CN	W	0	11	48	semiformell
CN	M	8,72	5,38	0,34	informell
CN	M	0	0	23,16	semiformell
CN	W	0	5	23	semiformell
CN	M	0	5	8,33	formell
DE	M	0	0,4	6,5	formell
DE	W	0	9,7	8,4	formell
DE	W	0	3,2	7,5	semiformell
DE	M	0	3,1	4,55	formell
DE	W	0	0,51	2,97	formell
DE	M	0	2,1	9,26	semiformell
DE	M	0	0	13,29	semiformell
DE	W	0	4,37	6,9	semiformell
DE	M	0,2	2,23	7,63	formell

DE	W	0	0	6,4	semiinformell
DE	M	0	0	12,5	semiinformell
DE	M	9,93	5,6	9,53	informell
DE	M	2,73	10,83	11,29	informell
DE	W	7,1	6,1	10,2	informell
EP	W	0,08	12,05	6,56	semiinformell
EP	M	1,8	7,24	9,34	semiinformell
EP	W	0,83	5,88	10	semiinformell
EP	M	0,04	7,48	16,28	formell
EP	M	7,67	19,06	9,28	formell
EP	M	2,06	6,2	5,4	semiinformell
EP	W	0,25	10,4	8,5	semiinformell
EP	M	3,98	8,52	11,75	semiinformell
EP	M	7,06	10,75	9,03	semiinformell
EP	W	0,7	6,7	6,98	semiinformell
EP	W	7,1	10,6	12,7	semiinformell
EP	W	0,24	7,79	17,77	semiinformell
EP	M	13,3	14,87	23,57	informell
ES	W	0			formell
ES	W	0,14	3,75	3,75	formell
ES	W	0,44	4,15	6,18	formell
ES	W	3,66	11,47	10,48	semiinformell
ES	W	1,85	6,78	9,4	semiinformell
ES	M	3,31	6,12	9,49	formell
ES	W	2,45	0	14,11	formell
ES	W	3,62	6,9	11,4	formell
ES	M	0,32	5,03	5,75	semiinformell
ES	W	0	14,25	10,3	formell
ES	M	15,51	5,39	8,36	informell
ES	W	14,12	4,11	10,59	informell

Tabelle 10.8: Einfaktorielle Varianzanalyse: F-Werte (Variablen: Häufigkeiten der Verwendung non-verbaler Kommunikationsmittel, Faktor: Sprache)

Nonverb. Mittel		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Gesten/min	Zw. d. Gruppen	760,32	4	190,08	11,534	0
	In d. Gruppen	1235,98	75	16,48		
	Gesamt	1996,3	79			
Gesichtsausdrücke/min	Zw. d. Gruppen	312,635	4	78,159	5,532	0,001
	In d. Gruppen	819,497	58	14,129		
	Gesamt	1132,131	62			
Bewegungen/min	Zw. d. Gruppen	1692,557	4	423,139	9,028	0
	In d. Gruppen	2718,425	58	46,869		
	Gesamt	4410,982	62			

Tabelle 10.9: Einfaktorielle Varianzanalyse: Post-Hoc-Vergleiche nach Scheffé (Variablen: Häufigkeiten der Verwendung nonverbaler Kommunikationsmittel, Faktor: Sprache)

Abhängige Variable			Mittlere Differenz	Standard fehler	Signifi kanz	95%-Konfidenzintervall	
						Untergrenze	Obergrenze
Gesten/min	BP	CN	9,40776*	1,4309	0	4,8886	13,927
		DE	8,74338*	1,63563	0	3,5776	13,9092
		EP	6,69989*	1,66308	0,005	1,4474	11,9524
		ES	6,38418*	1,69454	0,01	1,0323	11,736
	CN	BP	-9,40776*	1,4309	0	-13,927	-4,8886
		DE	-0,66438	1,31394	0,992	-4,8142	3,4854
		EP	-2,70787	1,34796	0,408	-6,9651	1,5494
		ES	-3,02357	1,38659	0,323	-7,4028	1,3557
	DE	BP	-8,74338*	1,63563	0	-13,9092	-3,5776
		CN	0,66438	1,31394	0,992	-3,4854	4,8142
		EP	-2,04349	1,56358	0,789	-6,9817	2,8948
		ES	-2,35919	1,59701	0,703	-7,403	2,6846
	EP	BP	-6,69989*	1,66308	0,005	-11,9524	-1,4474
		CN	2,70787	1,34796	0,408	-1,5494	6,9651
		DE	2,04349	1,56358	0,789	-2,8948	6,9817
		ES	-0,31571	1,62511	1	-5,4483	4,8169
	ES	BP	-6,38418*	1,69454	0,01	-11,736	-1,0323
		CN	3,02357	1,38659	0,323	-1,3557	7,4028
		DE	2,35919	1,59701	0,703	-2,6846	7,403
		EP	0,31571	1,62511	1	-4,8169	5,4483
Gesichtsausdrücke/min	BP	CN	2,37838	1,61161	0,704	-2,7492	7,5059
		DE	3,89393	1,66595	0,257	-1,4065	9,1944
		EP	-2,47827	1,68909	0,708	-7,8523	2,8958
		ES	1,15523	1,74661	0,979	-4,4018	6,7123
	CN	BP	-2,37838	1,61161	0,704	-7,5059	2,7492
		DE	1,51555	1,3566	0,869	-2,8007	5,8317
		EP	-4,85665*	1,38492	0,023	-9,263	-0,4504
		ES	-1,22316	1,45451	0,949	-5,8509	3,4046
	DE	BP	-3,89393	1,66595	0,257	-9,1944	1,4065
		CN	-1,51555	1,3566	0,869	-5,8317	2,8007
		EP	-6,37220*	1,44779	0,002	-10,9785	-1,7659
		ES	-2,7387	1,5145	0,519	-7,5573	2,0799
	EP	BP	2,47827	1,68909	0,708	-2,8958	7,8523
		CN	4,85665*	1,38492	0,023	0,4504	9,263
		DE	6,37220*	1,44779	0,002	1,7659	10,9785
		ES	3,6335	1,53992	0,248	-1,266	8,5329
	ES	BP	-1,15523	1,74661	0,979	-6,7123	4,4018
		CN	1,22316	1,45451	0,949	-3,4046	5,8509
		DE	2,7387	1,5145	0,519	-2,0799	7,5573
		EP	-3,6335	1,53992	0,248	-8,5329	1,266
Bewegungen/min	BP	CN	-13,08206*	2,93525	0,002	-22,4209	-3,7432
		DE	-0,86643	3,03422	0,999	-10,5202	8,7873
		EP	-3,835	3,07636	0,816	-13,6229	5,9529
		ES	-1,58864	3,18112	0,993	-11,7098	8,5325
	CN	BP	13,08206*	2,93525	0,002	3,7432	22,4209
		DE	12,21563*	2,4708	0	4,3545	20,0768
		EP	9,24706*	2,52237	0,015	1,2218	17,2723
		ES	11,49342*	2,64913	0,002	3,0649	19,922
	DE	BP	0,86643	3,03422	0,999	-8,7873	10,5202

		CN	-12,21563*	2,4708	0	-20,0768	-4,3545
		EP	-2,96857	2,63688	0,866	-11,3582	5,421
		ES	-0,72221	2,75838	0,999	-9,4984	8,0539
	EP	BP	3,835	3,07636	0,816	-5,9529	13,6229
		CN	-9,24706*	2,52237	0,015	-17,2723	-1,2218
		DE	2,96857	2,63688	0,866	-5,421	11,3582
		ES	2,24636	2,80467	0,957	-6,6771	11,1698